DOCUMENT RESUME

ED 386 459	TM 023 263

AUTHOR	Zahs, Daniel; And Others
TITLE	High School and Beyond Fourth Follow-Up Methodology Report. Technical Report.
INSTITUTION	National Opinion Research Center, Chicago, Ill.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	ISBN-0-16-045527-8; NCES-95-426
PUB DATE	Feb 95
NOTE	136p.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Records; Computer Oriented Programs; Data Analysis; *Data Collection; *Educational Attainment; *Followup Studies; Higher Education; High Schools; *High School Students; Longitudinal Studies; National Surveys; *Research Methodology; Telephone Surveys
IDENTIFIERS	*High School and Beyond (NCES)

ABSTRACT

This report describes and evaluates the methods, procedures, techniques, and activities that produced the fourth (1992) follow-up of the High School and Beyond (HS&B) study. HS&B began in 1980 as the successor to the National Longitudinal Study of the High School Class of 1972. The original collection techniques of HS&B were replaced by computer assisted telephone interviews, and other electronic techniques replaced the original methods. HS&B data are more user-friendly and less resource-dependent as a results of these changes. There were 2 components to the fourth follow-up: (1) the respondent survey which was a computer assisted telephone interview (CATI) based on 14,825 members of the 1980 sophomore cohort, and (2) a transcript study based on the 9,064 sophomore cohort members who reported postsecondary attendance. The response to the respondent survey was 85.3%. Response rate for the transcript study varied from 50.4% at private, for-profit institutions to 95.1% at public, four-year institutions. Technical innovations in this survey round included verification and correction of previously collected data through the CATI instrument, online coding applications, and statistical quality control. Survey data and information about the methodology are presented in 49 tables. An appendix contains the transcript request packages. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 386 459

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report

February 1995

High School and Beyond Fourth Follow-Up Methodology Report

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement

NCES 95-426

TM023263

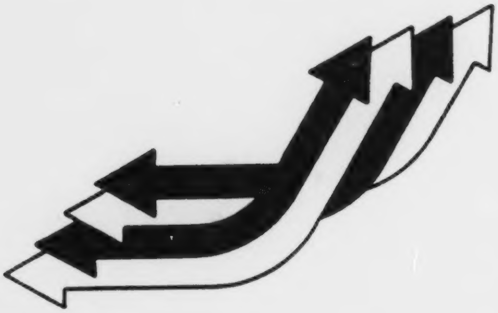
NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report **February 1995**

High School and Beyond Fourth Follow-Up Methodology Report

Daniel Zahs
Steven Pedlow
Marjorie Morrissey
Patricia Marnell
Bronwyn Nichols
National Opinion Research Center

C. Dennis Carroll
Project Officer
National Center for Education Statistics



U.S. Department of Education
Office of Educational Research and Improvement **NCES 95-426**

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, and analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

February 1995

Contact:
C. Dennis Carroll
(202) 219–1774

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-045527-8

Executive Summary

The High School & Beyond Fourth Follow-Up had two components: the respondent survey and the transcript study. The respondent survey was the fifth round of survey using computer assisted telephone interview (CATI) to survey a sample of 14,825 members of the 1980 sophomore cohort. The transcript study was based on the 9,064 sophomore cohort members who reported postsecondary attendance.

The issues addressed by the survey included:

- access to and choice of undergraduate and graduate educational institutions;
- persistence in attaining educational goals;
- progress through the curriculum;
- rates of degree attainment and other assessments of educational outcomes;
- barriers to persistence and attainment;
- rates of return to the individual and society; and
- relationship between course-taking patterns, academic achievement, and subsequent occupational choices and success

The field periods for data collection were as follows:

- CATI survey: February, 1992 to January, 1993
- Transcript study: December, 1992 to October, 1993

The CATI survey response rate was 85.3 percent and the average administration time was 30.6 minutes. The transcript study response rates varied by institution type from 50.4 percent at private, for-profit institutions to 95.1 percent at public, 4-year institutions. The response rate by students reporting postsecondary attendance was 93.2 percent (with at least one transcript). The transcript level response rate was 90.1 percent. Nonresponse was slightly higher for the fourth follow-up than previous rounds.

For both the CATI and the transcript study the estimated design effect (DEFF) was 2.0. This design effect is very similar to that for prior rounds.

Technical innovations used in this round included:

- verification and correction of previously collected data through the CATI instrument
- online coding applications that were used during interview and for coding transcripts
- statistical quality control

Foreword

This report describes and evaluates the methods, procedures, techniques, and activities that produced the fourth (1992) follow-up of the High School and Beyond (HS&B) study. HS&B began in 1980 as the successor to the National Longitudinal Study of the High School Class of 1972 (NLS-72). NLS-72 data spanned the period 1972 through 1986. HS&B now spans the period 1980-1992. Without a large increase in funding, both of these studies will not benefit from another follow-up. Hence, for HS&B, this report is the final documentation for this vast, rich dataset.

Over the years, HS&B matured. Paper and pencil collection techniques were replaced with computer assisted telephone interviews; hardcopy manuals were replaced with electronic codebooks; and mainframe computer tapes were replaced with personal computer compact disks. The HS&B data are more accurate, more user-friendly, and less-resource dependent as a result of these changes.

The National Center for Education Statistics (NCES) has been pleased to sponsor HS&B. NCES worked with the following U.S. Education Department offices that supplied supplementary funding: the Office of Bilingual and Minority Language Affairs, the Office for Vocational Education, the Office for Civil Rights, and the Office for Postsecondary Education. With funds from the Department of Defense, the National Science Foundation, and the Department of Health and Human Services, HS&B was further enhanced. Hopefully, the more than 600 articles, reports, papers, and dissertations based on HS&B will grow in number.

We hope that the information provided in this report will be helpful to HS&B users. We welcome comments for improving the format, content, and other aspects of this report and HS&B in general.

Paul D. Planchon
Associate Commissioner

Acknowledgments

The authors wish to thank the many people who contributed to the production of this report and High School and Beyond (HS&B). Project directors at the National Opinion Research Center have included Carol Stocking, Fan Calloway, Calvin Jones, Penny Sebring, Barbara Campbell, and Marjorie Morrissey. Project officers at the National Center for Education Statistics (NCES) have included Edith Huddleston, Bill Fetters, Sam Peng, Ricky Takai, Anne Weinheimer, Helen Ashwick, Jeff Owings, Roan Quintana-Garcia, Paula Knepper, and Dennis Carroll. Without the support of Marie Eldridge, Emerson Elliott, David Sweet, Sam Peng, Ron Hall, and Paul Planchon, HS&B would not have existed. Clifford Adelman deserves special thanks for applying his expertise to the transcript components of HS&B. Finally, special thanks are due the 58,000 students, 1,100 high schools, 7,000 parents, and 5,000 postsecondary institutions that provided the HS&B data.

This report was reviewed and improved by Jeff Owings, Nabeel Alsalam, and Bob Burton at NCES. Dan Madzellan of the Office of Postsecondary Education and Carol Fuller of the National Institute for Independent Colleges and Universities also kindly served as reviewers.

TABLE OF CONTENTS

Page

- 1. INTRODUCTION ... 1
  - 1.1 Overview ... 1
    - 1.1.1 NCES's Educational Longitudinal Studies Program ... 1
    - 1.1.2 High School and Beyond and NLS-72 ... 2
  - 1.2 History of High School and Beyond ... 2
    - 1.2.1 The Base Year Survey ... 2
    - 1.2.2 The First Follow-Up Survey ... 4
    - 1.2.3 The Second Follow-Up Survey ... 5
    - 1.2.4 The Third Follow-Up Survey ... 6
    - 1.2.5 The Fourth Follow-Up Survey ... 6
    - 1.2.6 Transcripts ... 7
  - 1.3 Related Studies and Data Files ... 8
    - 1.3.1 Base Year Files ... 8
    - 1.3.2 Other HS&B Files ... 8

- 2. STUDENT DATA COLLECTION INSTRUMENTS ... 11
  - 2.1 Base Year Survey ... 11
  - 2.2 First Follow-up Survey ... 12
    - 2.2.1 First Follow-up Sophomore Questionnaire ... 12
    - 2.2.2 1980 Sophomore Cohort (Not Currently In High School) Questionnaire ... 12
    - 2.2.3 Transfer Supplement ... 12
    - 2.2.4 Early Graduate Supplement ... 13
    - 2.2.5 First Follow-up Tests ... 13
  - 2.3 Second Follow-up Survey ... 13
  - 2.4 Third Follow-up Survey ... 14
  - 2.5 Fourth Follow-up Survey ... 14

- 3. SAMPLE DESIGN AND IMPLEMENTATION ... 17
  - 3.1 Base Year Survey Sample Design<1> ... 17
  - 3.2 First Follow-Up Survey Sample Design ... 19
  - 3.3 High School Transcripts Sample Design (1980 Sophomore Cohort) ... 20
  - 3.4 Second and Third Follow-Up Survey Sample Design ... 21
  - 3.5 Fourth Follow-Up Survey Sample and Transcript Study Design ... 24
  - 3.6 Sample Weights ... 24
    - 3.6.1 General Approach to Weighting ... 24
    - 3.6.2 Weighting Procedures ... 26
    - 3.6.3 Results of Weighting ... 28
  - 3.7 Nonresponse Analyses ... 29
    - 3.7.1 General Considerations ... 29
    - 3.7.2 Analysis of Follow-Up Survey Student Nonresponse Rates ... 31
      - 3.7.2.1 Fourth Follow-Up Student Nonresponse Rates: School Variables ... 33
      - 3.7.2.2 Fourth Follow-Up Survey Student Nonresponse Patterns: Student-Level Variables ... 35
      - 3.7.2.3 Summary of Nonresponse Analyses ... 38

TABLE OF CONTENTS

Page
3.8 Standard Errors and Design Effects 39
  3.8.1 Base Year and First Follow-Up 40
  3.8.2 Second Follow-Up 45
  3.8.3 Third Follow-Up 47
  3.8.4 Fourth Follow-Up 49
  3.8.5 Transcript Data Collection 53

4. DATA COLLECTION 61
4.1 Overview 61
4.2 Data and Materials Collected from Schools and Teachers 61
  4.2.1 School Questionnaires 61
  4.2.2 Teacher Comment Forms 62
  4.2.3 Course Offerings and Enrollments: Academic Year 1981-82 62
  4.2.4 Data Collection Procedure: Schools and Teachers 63
4.3 Student Data Collection 65
  4.3.1 Base Year Data Collection 65
  4.3.2 First Follow-up Data Collection: 1980 Sophomore Cohort 66
4.4 Collection of Student Transcripts 69
4.5 Second Follow-Up Data Collection: 1980 Sophomore Cohort 70
4.6 Third Follow-Up Data Collection: 1980 Sophomore Cohort 71
4.7 Fourth Follow-Up Data Collection: 1980 Sophomore Cohort 73

5. DATA CONTROL, PREPARATION AND PROCESSING 75
5.1 Base Year Procedures 75
5.2 First Follow-Up Procedures 75
  5.2.1 Shipping and Receiving Documents 75
  5.2.2 Editing and Coding 76
  5.2.3 Data Retrieval and Validation 77
5.3 Second Follow-Up Procedures 77
  5.3.1 Shipping and Receiving Documents 77
  5.3.2 Editing and Coding 78
  5.3.3 Data Retrieval and Validation 78
5.4 Third Follow-Up Procedures 79
  5.4.1 Shipping and Receiving Documents 79
  5.4.2 Coding and Computer Assisted Data Entry 79
  5.4.3 Data Retrieval and Validation 80
5.5 Fourth Follow-Up Data Control and Processing 81
  5.5.1 Computer-Assisted Telephone Interviewing (CATI) 81
  5.5.2 Case Delivery to Interviewers 81
  5.5.3 Telephone Number Management System 82
  5.5.4 On-Line Coding 83
  5.5.5 Postsecondary Institution (FICE) Coding 83
  5.5.6 Monitoring 84

TABLE OF CONTENTS

Page

5.6 Data Processing 84
  5.6.1 Maintenance of Longitudinal Locator Databases 85
  5.6.2 Receipt Control Procedures 85
  5.6.3 Optical Scanning 86
  5.6.4 Machine Editing 87
  5.6.5 Data File Preparation 89

6. SOPHOMORE COHORT POSTSECONDARY EDUCATION TRANSCRIPT STUDY 91
6.1 Scope of the Postsecondary Education Transcript Studies 91
6.2 Transcript Data Collection 91
6.3 Transcript Data Collection Objectives 92
6.4 Mailout of Transcript Request to Institutions 92
6.5 Data Collection Results 93
  6.5.1 The Institution-Level Response Rate 93
  6.5.2 Transcript-Level Response Rate 95
  6.5.3 Student-Level Data Collection Results 97
6.6 Data Preparation 97
  6.6.1 Data Preparation Objectives 97
  6.6.2 Data Organization 98
6.7 Computer Assisted Data Entry (CADE) and Coding 99
6.8 Data Quality Management 100
6.9 Data Processing 100

7.0 DATA QUALITY 103
7.1 Monitoring 103
7.2 Item Non-Response 105
7.3 Consistency Between Third Follow-up and Fourth Follow-up Responses 110
  7.3.1 Race/Ethnicity 111
  7.3.2 Marital Status 112
7.4 Proprietary Institution Non-response Issues 114
  7.4.1 Proprietary Respondents vs. Proprietary Non-respondents 114
  7.4.2 Proprietary School Students vs. Non-proprietary School Students 115

1. INTRODUCTION

The High School and Beyond (HS&B) Fourth Follow-up Survey is the fifth wave of the longitudinal study of the high school sophomore class of 1980. This round differed from previous follow-ups in that it focused exclusively on the sophomore class. During the spring and summer of 1992, young persons who had participated in the 1980 base year survey were administered a Computer Assisted Telephone Interview (CATI) and asked to detail their activities since the last round of data collection in 1986. In 1992, education and employment information from 1982-1986 was verified and corrected as needed, and transcripts were obtained for respondents who had attended postsecondary institutions.

This report summarizes and documents the major technical aspects of the fourth follow-up survey, and includes information on the survey instruments employed, sample design and implementation, and data collection and processing procedures used in the HS&B base year and four follow-up surveys.

1.1 Overview

1.1.1 NCES's Educational Longitudinal Studies Program

The mission of the National Center for Education Statistics (NCES) includes the responsibility to "collect and disseminate statistics and other data related to education in the United States" and to "conduct and publish reports on specific analyses of the meaning and significance of such statistics" (Education Amendments of 1974, Public Law 92-380, Title V, Section 501, amending Part A of the General Education Provisions Act).

Consistent with this mandate, NCES instituted the National Education Longitudinal Studies (NELS) program, whose general aim is to study longitudinally the educational, vocational, and personal development of young people, beginning with their elementary or high school years, and the personal, familial, social, institutional, and cultural factors that may affect that development.

The overall NELS program utilizes longitudinal, time-series data in two ways: a cohort is surveyed at regular intervals over a span of years, and comparable data are obtained from successive cohorts that permit studies of trends relevant to educational and career development and societal roles. Thus far, the NELS program consists of three major studies: the National Longitudinal Study of the High School Class of 1972 (NLS-72), High School and Beyond (HS&B) and the National Education Longitudinal Study of 1988 (NELS:88).

The first major study, NLS-72, began by collecting comprehensive base year survey data from approximately 19,000 high school seniors in the spring of 1972. The NLS-72 first follow-up survey added nearly 4,500 individuals in the original sample who did not participate in the base year survey. Three more follow-up surveys were conducted with the full sample in 1974, 1976, and 1979, using a combination of mail surveys and personal and telephone interviews. The fifth follow-up survey, with a subsample of about 15,000 individuals, took place during the spring of 1986.

The second major survey, HS&B, began in the spring of 1980 with the collection of base year questionnaire and test data on over 58,000 high school seniors and sophomores. The first follow-up survey was conducted in the spring of 1982, the second follow-up in the spring of 1984, the third follow-up in the spring of 1986, and the fourth follow-up in the spring of 1992.

The third major survey, NELS:88, began with a survey of eighth graders in 1988 and recently completed its second follow-up survey in 1992. The third follow-up survey is underway and is expected to continue through 1994.

1.1.2 High School and Beyond and NLS-72

High School and Beyond was designed to build on NLS-72 in three ways. First, the base year survey of HS&B included a 1980 cohort of high school seniors that was directly comparable to the 1972 cohort. Replication of selected 1972 student questionnaire items and test items made it possible to analyze changes subsequent to 1972 and their relationship to recent federal education policies and programs. Second, the introduction of the sophomore cohort provided data on the many critical educational and vocational choices made between the sophomore and senior years in high school, thus permitting a fuller understanding of the secondary school experience and how it affects students. Finally, HS&B expanded the NLS-72 focus by collecting data on a range of life cycle factors, such as family formation, labor force behavior, intellectual development, and social participation.

1.2 History of High School and Beyond

1.2.1 The Base Year Survey

The base year survey was conducted in the spring of 1980, and called for a highly stratified national probability sample of over 1,100 secondary schools as the first stage units of selection. At the second stage, 36 seniors and 36 sophomores were selected in each school (in schools with fewer than 36 students in either of these groups, all eligible students were included). Special efforts were made to identify sampled students who were twins or triplets so that their co-twins or co- triplets could be invited to participate in the study. (Data from nonsampled twins and triplets are not included in the student data files, but are available in a separate Twin Data File, which links questionnaire data from the base year and first follow-ups for sampled and nonsampled twins for special analyses.) Over 30,000 sophomores and 28,000 seniors enrolled in 1,015 public and private high schools across the country participated in the base year survey. (Detailed information about the samples can be found in the HS&B sample design report for the base year: Martin R. Frankel, Luane Kohnke, David Bunanno, and Roger Tourangeau, Sample Design Report, National Center for Education Statistics, 1981).

Certain types of schools were oversampled to make the study more useful for policy analyses. These included:

Public schools with high percentages of Hispanic students to ensure sufficient numbers of Cuban, Puerto Rican, and Mexican students for separate analyses;

Catholic schools with high percentages of minority students

Alternative public schools; and

Private schools with high-achieving students.

The Hispanic supplement to the sample was funded jointly by the Office of Bilingual Education and Minority Language Affairs (OBEMLA) and the Office for Civil Rights (OCR) within the Department of Education.

Survey instruments in the base year of HS&B included:

A sophomore questionnaire

A senior questionnaire

Student identification pages

A series of cognitive tests for each cohort

A school questionnaire

A teacher comment checklist

A parent questionnaire (mailed to a sample of parents from both cohorts)

The student questionnaires focused on individual and family background, high school experiences, work experiences, and plans for the future. The student identification pages included information that would be useful in locating the students for future follow-up surveys, as well as a series of items on the students' use of, proficiency in, and educational experiences with languages other than English. The cognitive tests measured verbal and quantitative abilities in both cohorts. In addition, the sophomore test battery included achievement measures in science, writing, and civics, while seniors were asked to respond to tests measuring abstract and nonverbal abilities. Of the 194 test items administered to the HS&B senior cohort in the base year, 86 percent were identical to items that had been given to the NLS-72 base year respondents.

School questionnaires, which were filled out by an official in each participating school, provided information about enrollment, staff, educational programs, facilities and services, dropout rates, and special programs for handicapped and disadvantaged students. The teacher comment checklist provided teacher observations on students participating in the survey. The parent questionnaire elicited information about the effects of family attitudes and financial planning on postsecondary educational goals.

1.2.2 The First Follow-Up Survey

The first follow-up sample consisted of about 30,000 1980 sophomores and 12,000 1980 seniors. It retained the multi-stage, stratified, and clustered design of the base year sample, and all students who had been selected for inclusion in the base year survey, whether or not they actually participated, had a chance of being included in the first follow-up survey. (Unequal probabilities were compensated by weighting.) NCES attempted to survey all 1980 sophomores (including base year nonrespondents) who were still enrolled in their original base year schools. Certain categories of 1980 sophomores (early graduates, dropouts and transfers) no longer enrolled in their original schools were subsampled and certain categories were sampled with certainty.

The data collected for sophomores included information on school, family, work experiences, educational and occupational aspirations, personal values, and test scores of sample participants. Students are also classified by high school status as of 1982 (i.e., dropout, same school, transfer, or early graduate). For the senior cohort, information concerning high school and postsecondary experiences and their experiences comprise the main focus.

The first follow-up survey also included all nonsampled co-twins and triplets who had been identified and surveyed during the base year, provided that the sampled twin or triplet was retained for the follow-up. However, nonsampled twins and triplets were not included in the probability sample and were not given weights; their data appear only on a separate Twin Data File. As in the base year survey, there was a Hispanic supplement in the first follow-up survey, again supported by OBEMLA and OCR. During the first follow-up information was again gathered from parents and school administrators.

A first follow-up school questionnaire was requested from all schools selected in the base year (including those schools that refused to participate), except schools that had no 1980 sophomores, schools that had closed, and schools that had merged with other schools in the sample. Schools not in the base year sample that had received en masse transfers of students from base year schools were contacted to complete a first follow-up school questionnaire and to arrange student survey activities. These schools were not considered to be part of the probability sample of secondary schools and were not given weights. However, survey data from these schools are included in the first follow-up School Data File, and are available for merging with first follow-up student data.

For the senior cohort, a self-administered mail-back questionnaire was the basic method of data collection. Approximately 12,000 packets containing survey questionnaires, instruction sheets, and incentive payment checks were sent to sample members during the first week of February 1982. Postcards with dual messages seeking a quick reply from nonrespondents and thanking early respondents for their cooperation were mailed during the third week following the initial mailout. Approximately 75 percent of the targeted senior cohort members completed and returned first follow-up questionnaires by mail. Two weeks later, those who still had not responded were called by trained telephone interviewers. An additional 19 percent completed the questionnaires through either in-person or telephone interviews. Respondents who completed the questionnaire by telephone were required to have a copy of the questionnaire in front of them while doing so in order to keep the survey experience as similar as possible to that of the mail questionnaires. Follow-up interviewing was halted in mid-July 1982 after a response rate of 94 percent had been obtained.

For the sophomore cohort, first follow-up data were collected through group administrations of questionnaires and tests. The sophomore group administrations were conducted in either the sampled students' high school or an appropriate location off- campus; the location depended on the survey member's school enrollment status during the data collection period (February through May 1982). Group administrations were scheduled off- campus for sample members who were no longer attending the sampled schools. These individuals (e.g., transfer students, dropouts, early graduates) were contacted by NORC Survey Representatives and brought together in small groups of two to six participants. The same survey administration procedures were followed for both types of group administration.

Subsequent to the first follow-up survey, high school transcripts were sought for a probability subsample of nearly 18,500 members of the 1980 sophomore cohort. The subsampling plan for the Transcript Study emphasized retaining members of subgroups who are especially relevant to education policy analysis. Compared to the base year and first follow-up surveys, the Transcript Study sample design further increased the overrepresentation of racial and ethnic minorities (especially for those with above average HS&B achievement test scores), students who attended private high schools, school dropouts, transfers, early graduates, and students whose parent participated in the base year Parents' Survey on financing postsecondary education.

1.2.3 The Second Follow-Up Survey

Conducted during the spring and summer of 1984, the second follow-up survey retained probability samples of about 15,000 1980 sophomores and 12,000 1980 seniors. The sample for the senior cohort was unchanged from that used for the first follow-up survey, while the sample for the sophomore cohort was selected from among the 18,500 cases selected in 1982 for the High School Transcripts study. The sample design for the sophomore cohort was modelled after that used for the first and subsequent follow-ups of the senior cohort, in that subgroups of special relevance to education policy formation (high school dropouts from the sophomore cohort, members of racial and ethnic minorities, those with data from the base year Parents Survey, those enrolled in postsecondary educational institutions, and so forth) were retained in the second follow-up with substantially higher probabilities than others. However, all individuals selected for the base year survey had a nonzero chance of retention in the second follow-up, regardless of whether they participated in the base year or first follow-up surveys.

As in prior survey rounds, the Office of Bilingual Education and Minority Language Affairs provided additional support for the Hispanic supplement to HS&B in order to increase the size of the Hispanic sample for special analyses.

For both seniors and sophomores, the data collected covered work experience, postsecondary schooling, earnings, periods of unemployment, and so forth. For both cohorts, data were collected through a self-administered mail-back questionnaire. Packets containing survey questionnaires, instruction sheets, and incentive payment checks were sent to sample members during the first week of February 1984. Two weeks later, postcards thanking respondents for their cooperation and requesting the cooperation of nonrespondents were mailed to all sample members. Two weeks after the cards were sent, trained telephone interviewers called those who had still not responded and urged them to do so. If this failed, interviews were conducted by telephone or in person. Survey design required both respondents interviewed over the telephone and those interviewed in person to have a copy of the questionnaire in front of them, in order to minimize bias due to the method of administration.

1.2.4 The Third Follow-Up Survey

The senior and sophomore cohort samples for the third follow-up survey were the same as those used for the second follow-up. Again, survey activities were initiated for all sample members-- except for 38 persons who were known to be deceased. (The nonsampled twins and triplets, however, were not surveyed during this wave.)

The questionnaires used during the 1986 third follow-up were the same for both the sophomore and senior cohorts. To maintain comparability with prior waves, many questions from previous follow-up surveys were repeated. Respondents were asked to update background information and to provide information about their work experience, unemployment history, education and other training, family information (including marriage patterns), income, and other experiences and opinions.

As in the second follow-up survey, data were collected through mail-back questionnaires; approximately 27,000 packets of survey materials were mailed to the last known addresses of the sample members. Contact procedures for nonrespondents remained unchanged from the previous rounds. Three weeks after the initial mail-out, respondents who had not returned their questionnaires were sent a postcard reminder. Two weeks after the cards were sent, trained telephone interviewers called to urge those who had still not responded. If this failed, interviews were conducted by telephone or in person. Approximately 66 percent of both samples mailed back their completed questionnaires; 5 percent of the seniors and 6 percent of the sophomores were interviewed in person; and about 16 percent of the seniors and 19 percent of the sophomores were interviewed by telephone. The survey design again required respondents interviewed by telephone or in person to use a copy of the questionnaire during the interview to minimize the bias due to method of administration. Follow-up interviewing resulted in a completion rate of 88 percent for the seniors and 91 percent for the sophomores.

A transcript study was conducted of third follow-up sophomore cohort respondents who reported attending postsecondary institutions. By 1987, when the study was conducted, these sample members had been out of high school for 5 years - long enough for many to attain vocational certificates, associate's degrees, and/or baccalaureate degrees.

1.2.5 The Fourth Follow-Up Survey

The fourth follow-up survey sought to obtain valuable information on issues of access to and choice of undergraduate and graduate educational institutions, persistence in obtaining educational goals, progress through the curriculum, rates of degree attainment and other assessments of educational outcomes, and rates of return to the individual and society. The fourth follow-up student interview emphasized these five issue areas pertinent to 1980 high school sophomores now in their middle twenties. And this study was particularly well suited to examine each of these themes because: (1) many items in prior rounds were related to these themes, thus providing a temporal context, and (2) the respondents' age placed them at a time when new information concerning these themes would provide invaluable insights into the effects of secondary and postsecondary education.

The fourth follow-up sample of the sophomore cohort contained the same 15,000 members as the second and third follow-up surveys, and attempts were made to contact all but 56 deceased sample members. By the end of the fourth follow-up, NORC identified an additional 99 deceased sample members, which brought the overall total of deceased sample members of the sophomore cohort to 155.

For the first time, a Computer Assisted Telephone Interview (CATI) was used to collect data. On February 5, 1992, a letter was sent to sample members describing the study and informing them that telephone interviewers would contact them to complete a telephone interview. The following week, telephone interviewing began.

Locating efforts occurred in both the phone center and in the field. Field interviewers were sent to locate respondents and encourage them to contact the telephone center in order to complete an interview. About 4,000 cases, or 28 percent, were located through the combined effort of the phone center and the field. Although 66.3 percent of the interviews were complete by September 19, locating and interviewing continued until the last week of January, 1993 when the study had reached a completion rate of 85.3 percent.

1.2.6 Transcripts

In 1993, another postsecondary transcript study was conducted to gather accurate and reliable data on the students' academic histories since leaving high school. Six years had passed between the third and fourth follow-up, allowing some sophomore cohort members to persist in obtaining their baccalaureate degrees and others to pursue graduate, doctoral, and first professional degrees (e.g., M.D., J.D.).

Because the fourth follow-up CATI instrument allowed interviewers to verify postsecondary attendance and to collect any new attendance information, those who completed their postsecondary schooling by 1987 were identified. If their transcripts were obtained during the 1987 transcript study, no request for transcripts was made in 1993. Instead, their transcript data were abstracted from the 1987 transcript files, recoded, and integrated with data from transcripts collected in 1993.

In February 1993, requests for transcripts were mailed to vocational and academic institutions for those sophomore cohort members who reported postsecondary attendance not covered by the 1987 transcript study. Prompting efforts began in the second week of April, when the completion rate was 47 percent. Including the 1987 transcript data, about 14,000 transcripts were processed from 15,000 institutions.

1.3 Related Studies and Data Files

In addition to the core surveys described above, records studies have been undertaken including the collection of the high school transcripts of the sophomore cohort and postsecondary education transcripts and financial aid data for the seniors. Data files for these studies and other HS&B data, such as parent surveys, school surveys, etc., are described below. These auxiliary data files greatly expand the core data sets potential and usefulness, and researchers are encouraged to become familiar with them.

1.3.1 Base Year Files

The Language File contains information on each student who during the base year reported some non-English language experience either during childhood or at the time of the survey. This file contains about 11,000 records (sophomores and seniors combined), with 42 variables for each student.

The Parent File contains questionnaire responses from the parents of about 3,600 sophomores and 3,600 seniors who are on the Student File. Each record on the Parent File contains a total of 307 variables, including parents' aspirations and plans for their children's postsecondary education.

The Twin and Sibling File contains base year responses from sampled twins and triplets, data on non-sampled twins and triplets of sample members, and data from siblings in the sample. This file (about 3,000 records) includes all of the variables that are on the HS&B student file, plus two additional variables (family ID and SETTYPE--type of twin or sibling).

The Sophomore Teacher Comment File contains responses from about 14,000 teachers on 18,000 students from 600 schools. The Senior Teacher Comment File contains responses from 14,000 teachers on 17,000 students from 600 schools. At each grade level, teachers had the opportunity to answer questions about HS&B sampled students who had been in their classes. The typical student in the sample was rated by an average of four different teachers. These files contain approximately 76,000 teacher observations of sophomores and about 67,000 teacher observations of seniors.

The Friends File contains identification numbers of students in the HS&B sample who were named as friends of other HS&B-sampled students. Each record contains the ID of sampled students and IDs of up to three friends, which can be used to trace friendship networks and to investigate the sociometry of friendship structures, including reciprocity of choices among students in the sample.

1.3.2 Other HS&B Files

The High School Transcript File describes the course-taking behavior of 16,000 sophomores of 1980 throughout their four years of high school. Data include a six-digit course number <1> for each course taken along with course credit, course grade, and year taken. Other items of information such as grade point average, days absent, and standardized test scores are also contained on the file.

The Offering and Enrollments File contains school information, course offerings, and enrollment data for about 1,000 schools. Each course offered by a school is identified by a six-digit course number. Other information such as credit offered by the school is also contained on each record.

The Updated School File contains base year data and first follow-up data from the 1,015 participating schools in the HS&B sample. First follow-up data were requested only from those schools that still existed in the spring of 1982 and had members of the 1980 sophomore cohort currently enrolled. Each high school is represented by a single record that includes 230 data elements from the base year school questionnaire, if available, along with other information from sampling files (e.g., stratum codes, case weights).

The Postsecondary Education Transcript File for the HS&B Seniors contains transcript data on dates of attendance, fields of study, degrees earned, and the titles, grades, and credits of every course attempted at each institution, coded into hierarchical files with the student as the highest level of aggregation. Although no survey forms were used, detailed procedures were developed to extract and process information from the postsecondary institution transcripts for all members of the 1980 senior cohort who reported attending any form of postsecondary schooling in the first or second follow-up surveys. (Over 7,000 individuals reported over 11,000 instances of postsecondary institution attendance.)

The Senior Financial Aid File contains financial aid records from respondents who reported attending postsecondary institution and federal records of the Guaranteed Student Loan Program and the Pell Grant program.

The Sophomore Financial Aid File contains information from federal records from the Guaranteed Student Loan program and from the Pell Grant program for all students who reported postsecondary education and who had participated in either of these two programs.

The HS&B HEGIS and PSVD File contains the postsecondary institution codes for schools HS&B respondents reported attending in the first and second follow-ups. In addition, the file provides data on institutional characteristics such as type of institution, highest degree offered, enrollment, admissions requirements, tuition, and so forth. This file permits analysts to link HS&B questionnaire data with institutional data for postsecondary institutions attended by respondents.

END NOTE

<1> Corresponds with descriptions in A Classification of Secondary School Courses (CSSC), Evaluation Technologies, Inc., July 1982.

2. STUDENT DATA COLLECTION INSTRUMENTS

Information on the 1980 sophomore cohort has come primarily from questionnaires filled out by students, school administrators, teachers, and parents of students. These data have been supplemented by information on courses taught at sampled schools, the number of students enrolled in those courses, and by information from students' high school transcripts. The survey instruments given to school officials, teachers, and parents, as well as the protocols and procedures governing the transmittal of information on course offerings and student transcripts, are described in the user's manuals for each of these data files created before the fourth follow-up. The base year senior and sophomore questionnaires were similar, with approximately three-fourths of the items in each version common to both. Features of the sophomore questionnaires used in the base year and subsequent follow-ups of High School and Beyond are described below.

2.1 Base Year Survey

Most of the questions in the sophomore questionnaire focused on students' behavior and experiences in the secondary school setting. Also included were questions about employment outside the school, postsecondary educational and occupational aspirations, and personal and family background. A small number of questions dealt with personal attitudes and beliefs. In addition, to facilitate the recontacting of students in later follow-up surveys, students were asked to provide complete addresses and telephone numbers for themselves and for some other person who would always know their whereabouts. Sophomores also completed a battery of cognitive tests which are described briefly below:

Vocabulary (21 items, 7 minutes): Used a synonym format.

Reading (20 items, 15 minutes): Consisted of short passages (100-200 words) followed by comprehension questions and a few analysis and interpretation items.

Mathematics (38 items, 21 minutes): Students were asked to determine which of two quantities was greater, whether they were equal, or whether there was insufficient data to answer the question.

Science (20 items, 10 minutes): Based on science knowledge and scientific reasoning ability.

Writing (17 items, 10 minutes): Based on writing ability and knowledge of basic grammar.

Civics Education (16 questions, 5 minutes): Based on various principles of law, government, and social behavior.

2.2 First Follow-up Survey

2.2.1 First Follow-up Sophomore Questionnaire

The first follow-up sophomore questionnaire documented secondary school experiences, especially shifts in attitudes and values since the base year, as well as work experiences and plans for postsecondary education. Almost all of the first follow-up questions had been asked in the base year; most were from the sophomore document, but many had appeared in the senior questionnaire only. Content areas in the sophomore questionnaire included education (high school program, courses taken, grades, standardized tests taken, attendance and disciplinary behavior, parental involvement, extracurricular and leisure activities, assessment of quality of school and teachers), postsecondary education (goals, expectations, plans, and financing), work/labor force participation (occupational goals, attitudes toward military service), demographics (parents' education, father's occupation, family composition, school age siblings, family income, marital status, race, ethnicity, sex, birthdate, physical handicaps), and values (attitudes toward life goals, feelings about self, and so forth).

Approximately 30 items in the sophomore questionnaire were identified as "critical" or "key" questions, and special efforts were taken to ensure that respondents did not omit these items.

2.2.2 1980 Sophomore Cohort (Not Currently In High School) Questionnaire

The questionnaire designed for persons who had dropped out of high school focused on the reasons for dropping out and its impact on their educational and career development. About a dozen of the items were developed especially for students who left school before completion; the remainder of the questionnaire was made up of items used either in the regular 1980 sophomore cohort questionnaire or the 1980 senior cohort instrument. Content areas included circumstances of leaving school (reasons for leaving, evaluation of decision, plans for obtaining high school diploma or equivalent), participation in training programs and other postsecondary education, work (labor force participation, detailed job history, aspirations, Armed Forces service), financial status (dependency, income), marital status (spouse's education, occupation, dependents), demographics (parents' education, father's occupation, race, sex, ethnicity, date of birth), and other personal characteristics (physical handicaps, values, feelings about self). Thirty items were designated as critical.

2.2.3 Transfer Supplement

The Transfer Supplement was completed by members of the sophomore cohort who had transferred out of the base year sample high school to another high school. The supplement was completed in addition to the regular First Follow-up Sophomore Questionnaire. Most of the items in the Transfer Supplement were new items (except a few that were taken from the school questionnaire). Content areas included reasons for transferring and for selecting a particular school, identification of school, school location, grade respondent was in when he or she transferred, entrance requirements, length of interruption in schooling (if any) and reason, type of school (general, specialized), size of student body, and grades. The supplement was brief, taking about 10 minutes to complete. There were four critical items.

2.2.4 Early Graduate Supplement

The Early Graduate Supplement was developed for members of the sophomore cohort who graduated from high school ahead of schedule. They completed this questionnaire in addition to the regular First Follow-up Sophomore Questionnaire. The Early Graduate Supplement documented reasons for and circumstances of early graduation, the adjustments required to finish early, and respondents' activities compared with those of other out-of-school survey members (i.e., dropouts, 1980 seniors.) Content areas included reasons for graduating early, when decision was made (what grade), persons involved in the decision, course adjustments required, school requirements, and postsecondary education and work experience (the questions for the last area were identical to those in the senior cohort instrument). This supplement took about 10 to 15 minutes to complete. Nine items were designated as critical.

2.2.5 First Follow-up Tests

The sophomore cohort completed the same tests as in the base year. For the early graduates, transfer students, and dropouts, group administration sessions were held so that they could complete questionnaires and tests as well. Where this was not possible, NORC mailed only the questionnaire to respondents.

2.3 Second Follow-up Survey

The Second Follow-up Sophomore Questionnaire included 71 questions clustered around nine major sections: background information, education, other training, military experience, work experience, periods unemployed, family information, income, and experiences and opinions. As could be expected, the information gathered differs substantially from that collected for the first follow-up. By this time the majority of respondents were out of high school and enrolled in postsecondary school, working, or looking for work.

The questionnaire asked for detailed information on schools attended after high school (up to three schools). Respondents indicated the kind of institution attended; hours per week spent in class; the degree, certificate, or diploma being sought; and requirements completed. Financial information included questions on tuition and fees and scholarships. Data were also gathered on financial aid from both parents to the respondent and any siblings.

The survey also obtained a work history, including occupation, industry, gross starting salary, gross income, hours worked per week, length of time without a job, length of time looking for work, job training and job satisfaction. Family information covered the spouse's occupation and education, date of marriage(s), number of children, and income and benefits received by both the respondent and spouse.

There were 36 questionnaire items designated as critical, and any respondents who omitted these items or who provided inconsistent data were telephoned to obtain the missing data or to resolve the inconsistencies.

2.4 Third Follow-up Survey

The Sophomore Cohort Third Follow-up Questionnaire was the same as that for the senior cohort. To maintain comparability with prior waves, many questions from previous follow-up surveys were repeated. Respondents were asked to update background information and to provide information about their work experience, unemployment history, education and other training, family information, income, and other experiences and opinions. Event history formats were used to obtain responses about jobs held, institutions attended, periods of unemployment, and marriage patterns. A few new items were added covering graduate degree programs and on alcohol consumption habits.

There were 37 items in the third follow-up survey that were designated as critical. Respondents were telephoned in order to obtain missing data or to resolve inconsistencies.

2.5 Fourth Follow-up Survey

Emphasis in the fourth follow-up instrument was placed on gathering current and verifying/correcting historical data on the education backgrounds and work experiences of the sophomore cohort. In the education section, the four areas of interest were: (1) undergraduate and graduate access and choice; (2) persistence; (3) progress through curriculum; and (4) attainment and outcome assessment. Data gathered on work experience focused primarily on the individual and societal advantages gained through the attainment of additional education. The work experience data, when added to information about work experiences collected during prior rounds of HS&B, gives a continuous record of the respondents' work and educational experience since the inception of the HS&B study.

Related to work experience were questions on income and assets that explored differences in short-term and long-term earnings between individuals who entered and completed their postsecondary education and those who did not finish high school, or did finish high school but did not attend a postsecondary institution. Other issue areas for which data were gathered include factors affecting participation in the political process and community affairs, and family formation patterns and its relevance to continuance in postsecondary education.

Previous rounds of HS&B relied extensively on self-completion questionnaires. During the fourth follow-up a Computer Assisted Telephone Interview (CATI) was used to collect data.

The CATI program used by NORC for the High School and Beyond fourth follow-up was AutoQuest. The CATI instrument provided the following features to the data collection effort:

Display of interviewer instructions, survey questions, and response categories, and on-line help screens

Display of multiple questions per screen

Question displays including text modified to reflect answers to prior questions or data from previous rounds

Response validity checking based on range, type, and comparison to previous answers

Entry of open-ended or verbatim text

Branching or skipping based on previous answers and/or on preloaded data

Capacity to suspend an interview and restart it at another time

Capacity to review and change a previous response

A system for scheduling respondents for interviews.

The instrument for HS&B fourth follow-up made innovative use of several of these features. For example, in order to present a more conversational style of interview, wherever possible related groups of questions were presented together on one screen. The effect was a more streamlined application. Also, response categories were frequently presented as point-and-shoot style menus rather than as lists of text with codes. Over 100 data items were preloaded from previous rounds and confirmed or corrected by respondents in the course of the interview.

The interview was implemented as two AutoQuest instruments. The small first instrument was used to locate and verify the identity of the respondent and collect contacting outcome codes, while the second instrument contained all survey questions. The two instruments were linked so that with a few key strokes an interviewer could move easily between them.

The primary advantage of this arrangement was one of performance.

The most frequently used instrument was the locating instrument, which could quickly display case information. The larger instrument was not accessed until the interviewer had actually contacted the respondent and had obtained the respondent's consent to proceed with the interview.

3. SAMPLE DESIGN AND IMPLEMENTATION

3.1 Base Year Survey Sample Design<1>

In the base year, students were selected using a two-stage, stratified probability sample design with schools as the first- stage units and students within schools as the second-stage units. Sampling rates for each stratum were set so as to select in each stratum the number of schools needed to satisfy study design criteria regarding minimum sample sizes for certain types of schools. As a result, some schools had a high probability of inclusion in the sample (in some cases, equal to 1.0), while others had a low probability of inclusion. The total number of schools selected for the sample was 1,122, from a frame of 24,725 schools with grades 10 or 12 or both.<2> Sampling strata and the number of schools selected in each are shown in Tables 3.1 and 3.2. Within each stratum schools were selected with probabilities proportional to the estimated enrollment in their tenth and twelfth grades. Within each school, 36 seniors and 36 sophomores were randomly selected. In those schools with fewer than 36 seniors or 36 sophomores, all eligible students were drawn in the sample.

Substitution was carried out for schools that refused to participate in the survey, but there was no substitution for students who refused, whose parents refused, or who were absent on Survey Day and make-up days.<3> Substitution for refusal schools occurred only within strata. In certain cases no substitution was possible because a school was the sole member of its stratum.

Table 3.1--High school and beyond base year school sample selections special strata (oversampled)

	Number
Alternative public	50
Cuban public	20*
Cuban Catholic	10*
Other Hispanic public	106*
High performance private	12
Other non-Catholic private (stratified by four census regions)	38
Black Catholic	30*
Total (oversampled)	266

*These schools were defined as those having 30 percent or more of enrollment from the indicated ethnic subgroup.

Table 3.2--High school and beyond base year school sample selections regular strata (not oversampled)

	Number
Regular Catholic (stratified by four census regions)	48
Regular public (stratified by nine census divisions; racial composition enrollment; central-city, suburban, rural)	808
Total (not oversampled)	856

The realization of the sample by stratum is shown in Tables 3.3 and 3.4. Although the sample design specified that students in all but the special strata would be selected with approximately equal probabilities, the probabilities are only roughly equal. In addition, the students in special strata were selected with higher probabilities, in some strata with extremely high probabilities. Moreover, the sample as realized did not equal the sample as drawn, creating further deviations from a self- weighting sample. Consequently, each school (and student) was assigned a weight equal to the number of schools (or students) in the universes they represented. Since each student's overall selection probability (hence weight) was further influenced by the sample design for the follow-up surveys, the derivation of student case weights is discussed below. Calculation of school weights is described in the High School and Beyond First Follow- up (1982) School Questionnaire Data File User's Manual.

Table 3.3-- High school and beyond base year sample realization, stage 1: sampling of schools

Stratum	Drawn in sample	Original schools*	Substituted schools	Total realized
Regular public	808	585	150	735
Alternative public	50	41	4	45
Cuban public	20	11	--	11
Other Hispanic public	106	72	30	102
Regular Catholic	48	40	5	45
Black Catholic	30	23	7	30
Cuban Catholic	10	7	2	9
High performance private	12	9	2	11
Other non-Catholic private	38	23	4	27
Total	1,122	811	204	1015

*Includes additional selections made when schools were found to be out-of scope.

Table 3.4-- High school and beyond base year sample realization, stage 2: sampling of students

	Total drawn in sample	Absent, both survey and make-up days	Student refused	Parent refused	Parental materials missing*	Total realized
Number	70,704	8,278	1,759	223	2174	58,270
Percent	100.0	11.7	2.5	0.3	3.1	82.4

*Unusable because of critical survey materials missing.

Use of weights should lead to correct estimates (within sampling error) of the population of 10th and 12th grade students in United States schools in spring 1980, and correct estimates of subgroups within it. Several analyses conducted since the base year survey have shown consistently that the weights give estimates reasonably close to those from other data sources.

3.2 First Follow-Up Survey Sample Design

The first follow-up sophomore and senior cohort samples were based on the High school and Beyond base year samples, retaining the essential features of a stratified multi-stage design; (for further details see Tourangeau, et al., 1983).<4> The important features of the first follow-up design were as follows.

For the sophomore cohort, all schools selected for the base year sample were contacted for participation in the first follow-up school survey except those that had no 1980 sophomores, had closed, or had merged with other schools in the sample. Schools that received two or more students from base year schools were included in survey activities, and school-level data from these institutions were eventually added to students' records as contextual information; however, these schools were not added to the existing probability sample of schools. Of the 1,015 schools that participated in the base year survey, a total of 40 were dropped from the first follow-up sample: 11 because they had no sophomores in the base year; 5 because they had merged with other schools already in the sample; 17 because they were junior high schools or schools that were closed, sending all their 1980 students to a single "target school;" and 7 because they had closed and sent their 1980 students to a large number of geographically dispersed schools. The 17 "target schools" that had received pools of base year students were included in survey activities but not added to the sample. Thus, 975 schools from the base year sample plus the additional 17 "target schools" were contacted for the first follow-up survey.

The sophomores still enrolled in their original base year schools were retained with certainty, since the base year clustered design made it relatively inexpensive to resurvey and retest them.

Sophomore cohort students no longer attending their original base year schools (e.g., dropouts, early graduates, and those who had transferred as individuals to a new school) were subsampled. Certain groups were retained with higher probabilities in order to support statistical research on such policy issues as excellence of education throughout the society, access to postsecondary education, and transition from school to the labor force.

Students who transferred as a class to a different school were considered to be still enrolled if their original school had been a junior high school, had closed, or had merged with another school. Students who had graduated early or had transferred as individuals to other schools were treated as school leavers for the purposes of sampling.

The 1980 sophomore cohort school leavers were selected with certainty or according to predesignated rates designed to produce approximately the number of completed cases needed for each of several different sample categories. School leavers who did not participate in the base year were given a selection probability of 0.1. Table 3.5 shows the number of currently enrolled students and school leavers in each major school stratum.

For the 1980 senior cohort, students selected for the base year sample had a known, non-zero chance of being selected for the first and all subsequent follow-up surveys. The first follow-up sample consisted of 11,995 selections from the base year probability sample. This total includes 11,500 selections from among the 28,240 base year participants and 495 selections from among the 6,741 base year nonparticipants. In addition, 204 non-sampled co-twins or triplets (not part of the probability sample) were included in the first follow-up sample, resulting in a total of 12,199 selections.

Table 3.5--Sample allocation for first follow-up of 1980 sophomore cohort

Original base year school stratum	Student status				
	Currently enrolled*	Drop-out	Transfer	Early graduate	Total
Regular public	18,684	1,932	796	493	21,905
Alternative public	672	184	58	39	953
Cuban public	220	52	17	30	319
Other Hispanic public	2,375	336	121	86	2,918
Regular Catholic	1,372	19	57	10	1,458
Black Catholic	780	32	128	11	951
Cuban Catholic	252	15	25	8	300
High performance private	336	0	15	4	355
Other non-Catholic private	459	31	73	15	578
Total	25,150	2,601	1,290	696	29,737

*Currently enrolled in base year (or other related) school.

3.3 High School Transcripts Sample Design (1980 Sophomore Cohort)

Subsequent to the first follow-up survey, high school transcripts were sought for a probability subsample of nearly 18,500 members of the 1980 sophomore cohort. The subsampling plan for the Transcript Study emphasized the retention of members of subgroups of special relevance for education policy analysis. Compared to the base year and first follow-up surveys, the Transcript Study sample design further increases the overrepresentation of racial and ethnic minorities (especially those with above average HS&B achievement test scores), students who attend private high schools, school dropouts, transfers and early graduates, and students whose parents participated in the base year Parent's Survey on financing postsecondary education.

Transcripts were collected and processed for nearly 16,000 members of the sophomore cohort. Transcript data can be merged with student questionnaire data files using the case identification numbers common between the two files. The Data File Users's Manual for the HS&B High School Transcripts Study contains a full description of the sample design and other features of the transcript study.

3.4 Second and Third Follow-Up Survey Sample Design

The sample for the second follow-up survey of the 1980 sophomore cohort was based upon the transcripts study design. A total of 14,825 cases were selected from among the 18,500 retained for the transcript study. As was the case for the senior cohort, the sophomore cohort second follow-up sample included disproportionate numbers of sample members from policy-relevant subpopulations (e.g., racial and ethnic minorities, students from private high schools, high school dropouts, students who planned to pursue some type of postsecondary schooling, and so on). Sample weights have been provided to compensate for differential selection probabilities and participation rates across all survey waves. Tables 3.6 through 3.9 present several alternative tabulations of the second follow-up sample of the sophomore cohort.<5> The members of the senior cohort selected into the second follow-up sample consisted exactly of those selected into the first follow-up sample. The third follow-up was the last one conducted for the senior cohort.

Table 3.6--1980 Sophomore cohort second and third follow-up sample distribution by race/ethnicity typology

Category	Population size N	% of total	Second follow-up n	% of total
Hispanic				
Cuban/Puerto Rican	89,674	2.4%	990	6.7%
High achievement	85,762	2.3%	886	6.0%
Other Hispanic	299,802	7.9%	1,375	9.3%
Asian Pacific Islander	46,835	1.2%	430	2.9%
Native American	48,418	1.3%	292	2.0%
Black				
High achievement	84,500	2.2%	741	5.0%
Other	375,185	9.9%	1,295	8.7%
High achievement/ low-SES Whites	69,759	1.8%	388	2.6%
All others	2,679,309	70.9%	8,428	56.8%
Total	3,779,288	100.0%	14,825	100.0%

NOTE: For this typology, sample members were assigned to ethnic or racial categories on a sequential or hierarchical basis. That is, individuals who reported Cuban or Puerto Rican origin or descent in either the base year or first follow-up were so classified in this typology. High achievement Hispanics were then classified among the remaining non-Cuban/non-Puerto Rican cases. (Since some Cubans and Puerto Ricans were also "high achievement," the total number of high-achievement Hispanics is larger than shown in this table. "Other Hispanics" were then classified from among all remaining cases not assigned to the two previous categories. This procedure was repeated sequentially for each remaining category in the table. The result is a distribution of mutually exclusive categories whose contents sum to the population or sample size. The distributions presented mask considerable overlap among groups within the sample (e.g., Blacks who are also Hispanic).

Table 3.7--1980 Sophomore cohort second, third, and fourth follow-up sample distribution by first follow-up student status indicator

Student status category	Population size N	Population size % of total	Second follow-up n	Second follow-up % of total
Currently (1982) enrolled	2,755,522	72.9	11,012	74.3
Dropout	512,439	13.6	2,584	17.4
Transfer	330,393	8.7	753	5.1
Early graduate	180,934	4.8	476	3.2
Total	3,779,288	100.0	14,825	100.0

Note: categories presented above result from screening of cases for the first follow-up survey. Dropouts who returned to complete diplomas have been flagged in the composite variable HSDIPLOM in the public release data files.

Table 3.8--1980 Sophomore cohort second, third, and fourth follow-up sample distribution by base year school type

Base year school type	Population size N	Population size % of total	Second follow-up n	Second follow-up % of total
Public	3,425,292	90.6	11,724	79.1
Catholic	229,106	6.1	2,704	18.2
Other private	124,890	3.3	397	2.7
Total	3,779,288	100.0	14,825	100.0

Table 3.9--1980 Sophomore cohort second, third, and fourth follow-up sample distribution by data availability

Student characteristic	Population size N	Population size % of total	Second follow-up n	Second follow-up % of total
Parent data available	364,011	9.6%	2,534	17.1%
Parent data and PSE plans or high achievement	175,791	4.7%	2,049	13.8%
High school transcript data	3,344,251	88.5%	13,024	87.9%
Twin data*	39,984	1.1%	163	1.1%

NOTE: Row categories in this table are not mutually exclusive.
*Sampled twins only. An additional 275 non-sampled, co-twins were included in the HS&B Transcripts Study. Approximately 140 non-sampled co-twins were retained in the second follow-up, yielding about 150 twin pairs.

3.5 Fourth Follow-Up Survey Sample and Transcript Study Design

The fourth follow-up is composed solely of members from the sophomore cohort. The members of the sophomore cohort selected into the fourth follow-up sample consisted exactly of those selected into the second and third follow-up sample. For any student who ever enrolled in postsecondary education, complete transcript information was requested from the institutions indicated by the student.

3.6 Sample Weights

3.6.1 General Approach to Weighting

The general purpose of weighting is to compensate for unequal selection probabilities and to adjust for nonresponse. The weights are based on the inverse of the selection probabilities at each stage of the sample selection process and on nonresponse adjustment factors computed within weighting cells. The fourth follow-up had two major components, the collection of survey data and the collection of postsecondary transcript data. Nonresponse occurred during both of these data collection phases. Weights were computed to account for nonresponse during either phase.

For the survey data, two weights were computed. The first weight (FU4WT) was computed for all fourth follow-up respondents. The second weight (PANEL5WT) was computed for all fourth follow-up respondents who also participated in the base year and first, second and third follow-up surveys.

First, a raw weight (RAWWT), unadjusted for nonresponse in any of the surveys, was calculated and included on the data file. The raw weight provides the basis for analysts to construct additional weights adjusted for the presence of any combination of data elements. Although caution should be used if the combination of data elements results in a sample with a high proportion of missing cases.

Two additional weights were computed to facilitate the use of the postsecondary transcript data. The collection of transcripts was based upon student reports of postsecondary attendance during either the third or fourth follow-up. A student may report attendance at more than one school.

The first transcript weight (PSEWT1) was computed for students where we obtained at least one requested (i.e. student reported) transcript. It is therefore possible for a student who was not a respondent in the fourth follow-up (FU4WT=0), but who was a respondent in the third follow-up, to have a non-zero value for PSEWT1.

The second transcript weight (PSEWT2) is more restrictive. It was designed to assign weights only to cases that were deemed to have complete data. Only students who responded during the fourth follow-up (and hence students for whom we have a complete report of postsecondary education attendance) and for whom we received all requested transcripts received a non-zero value for PSEWT2. For those who did not complete the fourth follow-up interview, complete transcripts may have been obtained in the 1987 transcript study, but since we cannot be certain they are complete, they have been given a weight of zero.

Table 3.10 describes these weights (and others that were calculated during previous waves) for the sophomore cohort. All of these weights, except the two postsecondary transcript weights, project to the population of about 3,781,000 high school sophomores of 1980. The transcript weights project to the sub-population of students (approximately 2,532,000) who have attended a postsecondary institution.

Table 3.10--Sample case weights for sophomore cohort, base year through fourth follow-up survey

Weight	Applies to cases with	Number of cases having non-zero weights
RAWWT	All follow-up selections	14,825
BYWT*	Base Year questionnaire data	13,749
FU1WT*	First Follow-up questionnaire data	14,102
FU2WT	Second Follow-up questionnaire data	13,682
PANELWT3	Base year, first follow-up and second follow-up questionnaire data	12,423
TESTWT2	Second follow-up questionnaire data, base year, and first follow-up test data	10,786
TRWT2	Second follow-up questionnaire data and H.S. Transcript Study data	12,142
FU3WT**	Third follow-up questionnaire data	13,481
PANELWT4**	Base year, first follow-up, second follow-up, and third follow-up questionnaire data	11,708
TESTWT3**	Third follow-up questionnaire data, base year,and first follow-up test data	14,392
FU4WT**	Fourth follow-up questionnaire data	12,795
PANEL5WT**	Base year, first follow-up, second follow-up, third follow-up, and fourth follow-up questionnaire data	10,594
PSEWT1**	At least one postsecondary transcript	8,447
PSEWT2**	All postsecondary transcripts and participation in fourth follow-up.	6,004

* These weights are not the same as those calculated during the base year or first follow-up survey, but are adjusted for retention in the second follow-up.

**These counts include deceased persons, who have been given a weight in order to keep the population totals consistent with those of the base year survey.

Note: TESTWT2 and TESTWT3 were constructed only for cases for whom sufficient test data were available to construct a meaningful composite score (TEST).

3.6.2 Weighting Procedures

The weighting procedures consisted of two basic steps. The first step was the calculation of a preliminary follow-up raw weight based on the inverse of the cumulative probabilities of selection for the base year sample and up through the fourth follow-up survey. The second step carried out the adjustment of this preliminary weight to compensate for "unit" nonresponse--that is, for noncompletion of an entire questionnaire or some combination of survey instruments. These steps are described in more detail below.

Step 1: Calculation of preliminary raw weights. The first step in weighting the sample was to develop raw weights that adjust for the unequal selection probabilities of students. This weight is based on the inverse of the selection probabilities at each stage of the sample selection process.

For the sophomore cohort, the sample selection process was as follows:

1) Selection of schools into the base year sample.

2) Selection of students into the base year sample from the selected schools.

3) Selection of students into the first follow-up sample given that they had been selected into the base year sample.

4) Selection of students into the high school transcript sample given that they had been selected into the base year and first follow-up samples.

5) Selection of students into the second follow-up sample given that they had been selected into the base year, first follow-up and transcript samples. All cases selected for the second follow-up were retained in the third and fourth follow-up samples.

Thus the raw or preliminary weight for a student is as follows:

preliminary weight= (1/P1hi) x (1/P2hij) x (1/P3k) x (1/P4k) x(1/P5k)

where

P1hi = the base year stage-one (school-level) selection probability for the ith school in the hth superstratum (see Frankel, et al; Sample Design Report, 1981, p.153)

P2hij = the base year stage-two (student-level) selection probability for the jth grade in the ith school of the hth superstratum (see Frankel, et al, 1981, p 154).

P3k = probability of selection (retention) into the first follow-up for students in the kth sampling category.

P4k = probability of selection (retention) into the high school transcript study for students in the kth sampling category.

P5k = probability of selection (retention) into the second follow-up for students in the kth sampling category.

P1hi, the base year stage-one probability of selection, had been calculated during the base year and includes adjustments for ineligible and noncooperating schools. P2hij, the base year probability of selection for each student within his or her school and grade (given that the school had been selected), had been calculated during the base year as equal to the number of students selected in a grade within a school divided by the total number of students in that grade in the school. The values of P3k, P4k, and P5k, the probabilities of selection (retention) in the first follow-up, transcript study and second follow-up , depend on the specific sampling category in which a student was placed. These retention rates ranged from 1.0 for students retained with certainty to 0.1 for out-of-school base year nonparticipants.

Step 2: Nonresponse adjustment. In this step, the raw weights obtained in step 1 were multiplied by nonresponse ratio adjustment factors. As described earlier, different factors were used to develop FU4WT, PANELWT5, PSEWT1, and PSEWT2 but the approach is similar for each weight. Cases were distributed among weighting cells. Within each weighting cell two sums of raw weights were computed: the first, for all cases in the cell selected for the survey wave or combination of waves (selections); the second, for all cases in the cell for whom the specified combination of questionnaire and/or transcript data were collected (participants). The ratio of the two sums (selections over participants) provided a factor used to expand the preliminary weight of each participant to compensate for the missing weights of those who were selected but did not participate. The raw weights of nonparticipants were multiplied by an adjustment factor of zero to produce final weights of zero for these cases. Thus, the nonresponse adjustment consists of distributing the preliminary weights of the nonparticipants proportionately among the participants in each weighting cell.

The weighting cells were defined by cross-classifying cases by several variables. For the fourth follow-up weight (FU4WT), the cells were defined by:

(1) Dropout Status (as of Second Follow-Up) [HSDIPLOM]
    (1) non-dropout (diploma or GED obtained)
    (2) dropout

(2) School type (for non-dropouts only) [SCHSAMP]
    (1) regular public and alternative
    (2) .... Hispanic public
    (3) .... Catholic
    (4) .... private non-Catholic

(3) Sex [SEX]
    (1) male
    (2) female

(4) Race [RACE2]
  - (1) Hispanic
  - (2) non-Hispanic Black
  - (3) non-Hispanic White and other

(5) Base year test quartile [BYTESTQ]

for non-dropouts:	for dropouts:
(0) no test data available	(0) no test data available
(1) lowest quartile	
(2) second quartile	(1) below median
(3) third quartile	(2) above median
(4) highest quartile	

In some instances, cells were combined by pooling cases across base year test quartile classifications or type of high school attended. During the third and fourth follow-up, weights were generated for the deceased in order to more accurately determine the nonresponse adjustment and to permit analysis of prior survey data for these respondents.

3.6.3 Results of Weighting

As a check on the adequacy of the sample case weights, NORC analyzed the statistical properties of the weights. Table 3.11 shows the mean, variance, standard deviation, coefficient of variation, minimum, maximum, skewness, and kurtosis for each of the weights calculated for the fourth follow-up survey.

Table 3.11--Statistical properties of sample weights

	RAWWT	FU4WT	PANELWT5	PSEWT1	PSEWT2
Mean	255.0	295.5	356.9	299.7	421.7
Variance	57,703	77,638	96,542	73,782	140,146
Standard deviation	240.2	278.6	310.7	271.6	374.4
Coefficient of variation	0.942	0.943	0.871	0.906	0.888
Minimum	1.45	1.45	2.05	1.45	2.23
Maximum	3098	3465	4275	3176	4238
Skewness	2.38	2.72	2.01	2.03	1.92
Kurtosis	11.9	15.26	10.31	10.25	8.85
Number of cases	14825	12,795	10,594	8,447	6,004

3.7 Nonresponse Analyses

3.7.1 General Considerations

Nonresponse inevitably introduces some degree of error into survey results. In examining the impact of nonresponse, it is useful to think of the survey population as including two strata--a respondent stratum that consists of all units that would have provided data had they been selected for the survey, and a nonrespondent stratum that consists of all units that would not have provided data had they been selected. The actual sample of respondents necessarily consists entirely of units from the respondent stratum. Thus, sample statistics can serve as unbiased estimates only for the respondent stratum; as estimates for the entire population, the sample statistics will be biased to the extent that the characteristics of the respondents differ from those of the entire population.<6>

In the High School and Beyond study, there were two stages of sample selection and therefore two stages of nonresponse. During the base year survey, sample schools were asked to permit the selection of individual sophomores and seniors from school rosters and to designate "survey days" for the collection of student questionnaire and test data. Schools that refused to cooperate in either of these activities were dropped from the sample. Individual students at cooperating schools could also fail to take part in the base year survey. Unlike "refusal" schools, nonparticipating students were not dropped from the sample; they remained eligible for selection into the follow-up samples.

Estimates based on student data from the base year surveys include two components of nonresponse bias: bias introduced by nonresponse at the school level, and bias introduced by nonresponse on the part of students attending cooperating schools. Each component of the overall bias depends on two factors--the level of nonresponse and the difference between respondents and nonrespondents:

$$\text{Bias} = P1(Y1R - Y1NR) + P2(Y2R - Y2NR)$$

in which

P1 = the proportion of the population of students attending schools that would have been nonrespondents,

Y1NR = the parameter describing the population of students attending nonrespondent schools,

P2 = the proportion of students attending respondent schools who would have been nonrespondents, and

Y2NR = the parameter describing this group of students.

Nonresponse bias will be small if the nonrespondent strata constitute only a small portion of the survey population or if the differences between respondents and nonrespondents are small. The proportions P1 and P2 can generally be estimated from survey data using appropriately weighted nonresponse rates.

The implications of the equation can be easily seen in terms of a particular base year estimate. On the average, sophomores got 10.9 items right on a standardized vocabulary test.<7> This figure is an estimate of Y2R, the population mean for all participating students at cooperating schools. Now, suppose that sophomores at cooperating schools average two more correct than sophomores attending refusal schools (Y1R - Y1NR = 2), and suppose further that among sophomores attending cooperating schools, student respondents average one more correct answer than student nonrespondents (Y2R - Y2NR = 1). Noting that the base year school nonresponse rate was about .30 <8> and the student nonresponse rate for sophomores was about .12 <9>, we can use these figures as estimates of P1 and P2 and we can use this equation to calculate the bias as:

$$\text{Bias} = .30(2) + .12(1) = .72$$

That is, the sample estimate is biased by about .7 of a test score point.

This example assumes knowledge of the relevant population means; in practice, of course, they are not known and, although Pl and P2 can generally be estimated from the nonresponse rates, the lack of survey data for nonrespondents prevents the estimation of the nonresponse bias. The High School and Beyond study is an exception to this general rule: during the first follow-up, school questionnaire data were obtained from most of the base year refusal schools, and student data were obtained from most of the base year student nonrespondents selected for the first follow-up sample. These data provide a basis for assessing the magnitude of nonresponse bias in base year estimates.

The bias introduced by base year school-level refusals is of particular concern since it carries over into successive rounds of the survey. Students attending refusal schools were not sampled during the base year and have no chance for selection into subsequent rounds of observation. To the extent that these students differ from students from cooperating schools during later waves of the study, the bias introduced by base year school nonresponse will persist. Student nonresponse is not carried over in this way since student nonrespondents remain eligible for sampling in later waves of the study.

The results of three types of analyses concerning nonresponse are described in an earlier report.<10> Based on school questionnaire data, schools that participated during the base year were compared with all eligible schools. Based on the first follow-up student data, base year student respondents were compared with nonrespondents. Finally, student nonresponse during the first follow-up survey was analyzed. Taken together, these earlier analyses indicated that nonresponse had little effect on base year and first follow-up estimates. The results presented there suggest that the school-level component of the bias affected base year estimates by 2 percent or less and that the student-level component had even less impact.

3.7.2 Analysis of Follow-Up Survey Student Nonresponse Rates

This section examines the antecedents and correlates of nonresponse. A few preliminary remarks on the bias resulting from nonresponse are nonetheless in order. First, it should be noted that school nonresponse may have the same effect on base year, first, second, third, and fourth follow-up estimates-- students attending refusal schools were not sampled in the base year and have no chance of inclusion in the first, second, third, or fourth follow-up. For this reason, the estimates presented in earlier reports <11> may serve as estimates of the bias due to school nonresponse for the follow-up surveys as well as the base year. To the extent that the association between school attended and student characteristics decreases with the passage of time since the base year, the biasing effect of school refusals may be less now than it was for the base year. Student nonresponse was a little higher in the fourth follow-up than in the base year survey. Overall, the weighted student nonresponse rate during the fourth follow-up was 13.9 percent versus 12.0 percent during the base year. Thus bias in fourth follow-up estimates due to student nonresponse may be slightly larger than that in the base year estimates. However, bias in the base year was judged to be small.

Student nonresponse

There were several causes of student nonparticipation in the follow-up surveys. Some students refused to cooperate; others could not be located or were unavailable at the time of the follow-up surveys, and a few had died. Nonresponse rates were calculated in the usual way; the nonresponse rate is the proportion of the selected students (excluding deceased students) who were nonrespondents:

$$P = NR / (R + NR)$$

in which

P = the nonresponse rate
R = the number of responding students
NR = the number of nonresponding students.

Nonresponse rates were calculated by school-level and student-level variables using both unweighted and weighted data. The weight used was RAWWT. (See section 3.6 for a description of the weighting procedures.)

An overall indication of the level of participation and nonparticipation in the base year, first follow-up, second follow-up, third follow-up, and fourth follow-up surveys is presented in Table 3.12. This tables shows frequencies and percentages of cases in each of thirty-two cells. The totals presented in Table 3.12 are unweighted.

Table 3.12--Participation patterns for base year, first follow-up, second follow-up, third follow-up and fourth follow-up surveys

BY	1FU	2FU	3FU	4FU	Frequency	(Unwtd) Percent
N	N	N	N	N	63	0.4
N	N	N	N	Y	13	0.1
N	N	N	Y	N	7	0.0
N	N	N	Y	Y	16	0.1
N	N	Y	N	N	7	0.0
N	N	Y	N	Y	2	0.0
N	N	Y	Y	N	4	0.0
N	N	Y	Y	Y	14	0.1
N	Y	N	N	N	31	0.2
N	Y	N	N	Y	21	0.1
N	Y	N	Y	N	20	0.1
N	Y	N	Y	Y	40	0.3
N	Y	Y	N	N	24	0.2
N	Y	Y	N	Y	52	0.4
N	Y	Y	Y	N	114	0.8
N	Y	Y	Y	Y	637	4.3
Y	N	N	N	N	82	0.6
Y	N	N	N	Y	21	0.1
Y	N	N	Y	N	31	0.2
Y	N	N	Y	Y	59	0.4
Y	N	Y	N	N	20	0.1
Y	N	Y	N	Y	31	0.2
Y	N	Y	Y	N	48	0.3
Y	N	Y	Y	Y	292	2.0
Y	Y	N	N	N	140	1.0
Y	Y	N	N	Y	114	0.8
Y	Y	N	Y	N	106	0.7
Y	Y	N	Y	Y	334	2.3
Y	Y	Y	N	N	244	1.7
Y	Y	Y	N	Y	464	3.2
Y	Y	Y	Y	N	1089	7.4
Y	Y	Y	Y	Y	10530	71.8
Total					14670	100.0

NOTE: Counts refer to main samples only, excluding nonsampled co-twins, and excluding deceased persons.
BY = base year survey; 1FU = first follow-up survey;
2FU = second follow-up survey; 3FU = third follow-up survey;
4FU = fourth follow-up survey;
Y denotes participation, and N denotes non-participation

3.7.2.1 Fourth Follow-Up Student Nonresponse Rates: School Variables

This section examines nonresponse to the fourth follow-up by school-level variables. Five variables are shown in Table 3.13: school type, census region, level of urbanization, percentage of Black enrollment, and average enrollment. Base year and first follow-up data were used to classify the schools. The response rates given in the table are weighted, using RAWWT.

Students from alternative public schools had the highest nonresponse rate (24.3 percent) for all school types. Hispanic public school students were next highest (17.0 percent). Regular public and non-Catholic private appear somewhat similar (13.9 percent and 13.1 percent respectively). Students from Catholic schools had the lowest nonresponse rate (11.1 percent). There is some variation in nonresponse by region. The highest nonresponse rates occur in the West (16.3 percent) and the Northeast (16.1 percent). The lowest nonresponse rates occur among participants who had been students with the North Central region (10.9 percent). In regards to degree of urbanization, a pattern is seen. The higher the degree of urbanization the greater the degree of nonresponse. Students selected at schools with a large percentage of Blacks (25 percent or more) showed some what higher rates of nonresponse than students at schools with fewer Blacks. Student nonresponse seems to increase roughly with school size.

Table 3.13--Weighted student nonresponse rates by selected school characteristics

	Nonresponse rate (Percent)
Total population	13.9
School type:	
Regular public	13.9
Hispanic public	17.0
Alternative public	24.3
Non-Catholic private	13.1
Catholic	11.1
Region:	
Northeast	16.1
North Central	10.9
South	13.5
West	16.3
Urbanization:	
Urban	19.6
Suburban	13.6
Rural	10.1
Percent Black:	
25% or less	12.3
Greater than 25%	19.9
Other/unknown	14.0
Average enrollment:	
100 or less	11.2
101-325	11.2
326-550	13.4
More than 550	18.2
Other/unknown	15.7

Note: Deceased respondents (155 unweighted cases) have been excluded from both the numerator and denominator for the calculation of these nonresponse rates.

3.7.2.2 Fourth Follow-Up Survey Student Nonresponse Patterns: Student-Level Variables

In this section, the student nonresponse rates to the fourth follow-up survey are analyzed by student-level variables, including demographic characteristics, academic aptitude, high school program, and postsecondary education. Students were classified by their responses to the base year questionnaire for all characteristics except race and student status. For classification by race, first follow-up and base year data were used; for student status, first and second follow-up data were used. Table 3.14 shows the weighted rate of nonresponse by race, sex, high school academic program, base year SES, and student status. The category "other/unknown" is a general classification that includes both cases with missing data and cases that did not fall into any of the other specifically defined categories. Nonresponse generally is substantially higher for the "other/unknown" categories, because many sample members who were nonparticipants in earlier rounds, from which these variables were derived, were also nonparticipants in the fourth follow-up.

Table 3.14--Weighted student nonresponse rates by selected student characteristics

	Nonresponse rate (percent)
Total population	13.9
Race*:	
White	10.1
Black	20.6
Hispanic	19.4
Other/unknown***	40.3
Sex:	
Male	15.4
Female	12.3
Academic Program:	
General	16.0
Academic	10.2
Vocational	13.1
Other/unknown***	59.6
SES Quartile in base year:	
Highest quartile	9.5
Middle two quartile	11.7
Lowest quartile	14.7
Other/unknown***	38.8
Student Status**:	
No postsecondary education	14.3
Only vocational postsecondary education	9.2
Other postsecondary education	8.1
Unknown/missing***	46.2

Note: Deceased respondents (155 unweighted cases) have been excluded from both the numerator and denominator for the calculation of these nonresponse rates.
*Based on base year and first follow-up data.
** Based on base year, first follow-up and second follow-up data.
***"other/unknown" includes cases with missing data and cases who did not otherwise fall into any of the defined categories.

There is some variation in student nonresponse by race. Blacks and Hispanics show similar rates of nonresponse (20.6 and 19.4) with whites having a nonresponse rate at about half this level (10.1). Males exhibit a slightly higher nonresponse rate than females (the difference being slightly over 3%). Students who were in academic programs during the base year were less likely to be nonrespondents than students in general or vocational programs. Students classified within the highest level of SES showed the lowest level of nonresponse. Nonresponse increased as SES classification decreased. Students who had no postsecondary education (by the time of the second follow-up) had higher rates of nonresponse (14.3) than students with only vocational postsecondary education (9.2) or other postsecondary education (8.1).

These differences across groups in response rates are somewhat similar to those observed during previous rounds of data collection. A picture of student nonrespondents continues to emerge from the analyses; it suggests that groups with less involvement with education were less likely to participate in the survey. Dropouts had higher nonresponse rates than non-dropouts; students with lower grades and lower test scores showed higher nonresponse than students with higher grades and test scores; students who were frequently absent from school showed higher nonresponse than students absent infrequently; and students in vocational or general programs were more likely to be nonrespondents than students in academic programs.

3.7.2.3 Summary of Nonresponse Analyses

The analyses presented here support four general conclusions:

(1) The school-level bias component in estimates is small, averaging less than 2 percent for base year and first follow-up estimates. It is probably of a similar magnitude for fourth follow-up estimates.

(2) The student-level bias component in base year estimates is also small, averaging about 0.5 percent for percentage estimates.

(3) The student-level bias component in first, second, and third follow-up estimates is limited by the nonresponse rates, which were about three-fourths of the base year rates.

(4) The student-level bias component in the fourth follow-up is limited by the nonresponse rate, which was slightly higher than the base year rate.

The first and second conclusion together suggest that nonresponse bias is not a major contributor to error in base year estimates. The first and third suggest that nonresponse bias is not a major contributor to error in the first, second and third follow-up estimates either. The first and fourth conclusion suggest that nonresponse bias might be a little greater than for the previous follow-ups, but probably not by much.

Each of these conclusions must be given some qualifications. The analysis of school-level nonresponse is based on data concerning the schools, not the students attending them. The analyses of student nonresponse are based on survey data and are themselves subject to nonresponse bias. Despite these limitations, the results consistently indicate that nonresponse had a small impact on base year and follow-up estimates.

Nonresponse relating to the transcript study is discussed in Chapter 6.

3.8 Standard Errors and Design Effects

Statistical estimates calculated using High School and Beyond survey data are subject to sampling variability. Because the sample design for the HS&B cohorts involved stratification, disproportionate sampling of certain strata, and clustered (i.e., multi-stage) probability sampling, the calculation of exact standard errors for survey estimates can be difficult and expensive. Popular statistical analysis packages such as SPSS (Statistical Package for the Social Sciences) or SAS (Statistical Analysis System) normally calculate standard errors using the assumption that the data being analyzed were collected from simple random samples. The HS&B sample design is somewhat less efficient than simple random samples of equal size. Thus, sampling errors generated by SPSS and SAS under the assumption of simple random samples will often significantly underestimate the sampling variability of statistical estimates such as population means, percentages, and more complex statistics like correlations and regression coefficients.

Several procedures are available for calculating precise estimates of sampling errors for complex samples. Kish and Frankel <12> distinguish three major approaches to the computation of standard errors for statistics based on complex designs: Taylor Series approximations, Balanced Repeated Replication (BRR), and Jackknife Repeated Replication (JRR). These procedures vary somewhat not only in computational convenience and cost, but also in their ability to account for several sources of sampling variability, most notably clustered selection of sample cases. Sampling error estimates for the first and second follow-ups were calculated by the method of Balanced Repeated Replication (BRR), using BRRVAR, a Department of Education statistical subroutine called as a SAS procedure. Unfortunately, BRRVAR is no longer compatible with SAS. The BRR programs WESVAR and SUREG are now available commercially. For the base year, third and fourth follow-ups, Taylor Series approximations have been employed. More detailed discussions of the BRR and Taylor Series procedures can be found in the High School and Beyond Third Follow-Up Sample Design Report.

The Data Analysis System (DAS) that is included as part of the public release file, automatically reports design corrected Taylor-series standard errors for the tables it generates. Users of the DAS therefore need make no adjustments to these estimates. However, other users may wish to use other software when analyzing data from a restricted use file. Unfortunately, not all these users will have access to the programs needed to estimate standard errors for complex surveys. Thus, it is often useful to report design effects (DEFFs) in addition to standard errors for complex surveys such as the High School and Beyond survey. The design effect is a measure of how different the actual standard errors are from those that would be calculated under a simple random sample assumption with the same sample size. The square root of the design effect, called the root design effect (DEFT), is also useful, and both are defined as below:

DEFF . . . . =
VARest/VARsrs

48

and
DEFT . . . . =
square root of (VARest/VARsrs)
= . . . . SEest/SEsrs
in which

VARest = ... the actual variance of a sample estimate
VARsrs =
the estimate of variance that would be obtained if the sample were treated as a simple random sample
SEest = ... the actual standard error of a sample estimate
SEsrs = the estimate of variance that would be obtained if the sample were treated as a simple random sample

While design effects cannot be calculated for every estimate that users will be interested in, design effects will be similar from item to item within the same subgroup or population. In Tables 3.15-3.19, we calculated design effects for 30 items at each survey wave. Users can calculate approximate standard error estimates for items not in these tables by multiplying the standard error under the simple random sample assumption by the mean root design effect (DEFT) for the population being studied. The standard deviation of the root design effects in the tables give some indication of how close the mean root design effect is likely to be to the actual root design effect of the estimate.

For example, the simple random sample variance for proportions is just

$$= p(1 - p)/n$$

in which
p = the estimated proportion
n = the number of cases with non-missing data

The standard error of a proportion can then be estimated by multiplying the square root of the expression in the above equation by the mean root design effect (DEFT):

$$SE = DEFT \times SQRT\{(p[1 - p]/n)\}$$

3.8.1 Base Year and First Follow-Up

Table 3.15 displays standard errors and design effects for 30 proportions and seven averages based on weighted data from the first follow-up questionnaires and test. The mean root design effect for the 37 statistics is 1.8. This is somewhat higher than the mean (1.7) observed during the base year survey (see Frankel, et al; p. A-4). The sample of sophomores for the first follow-up differs from the base year sophomore sample in several key respects. Although the bulk of the base year sophomore sample was retained for the first follow-up with certainty, a few groups were subsampled. The subsampling introduces additional variability into the follow-up weights; the added variability of the weights reduces the efficiency of the sample, which is reflected in the larger design effects. The largest contributors to this loss of efficiency were base year nonparticipants who dropped out of school prior to the first follow-up. This group--consisting of about 500 selected cases--was sampled at a rate of .10; the mean follow-up weight for this group is about 15 times larger than the mean weight for the rest of the sample.

Table 3.16 displays estimates for the base year sophomore sample using data from base year participants who were selected for the first follow-up sample. The questionnaire items in table 3.16 are identical to those in table 3.15, but the estimated proportions and standard errors are based on responses to these items in the base year sophomore questionnaire. For the most part, these items were repeated verbatim in the first follow-up questionnaire. In one case, response options were reordered in the follow-up questionnaire. Table 3.16 shows that the mean DEFT is estimated to be 1.643; this is very close to the figure (1.651) calculated during the base year (see Frankel, et al; p.A-4). The mean DEFT in table 3.16 is lower than the mean in table 3.15 (1.6 vs. 1.8); the estimates for the follow-up sophomore sample are relatively less efficient than estimates for the base year sophomores. This difference probably reflects the increased variability of the follow-up weights as described above.

Table 3.15--Standard errors and design effects associated with estimated proportions and averages of first follow-up sophomores who had specified characteristics, using first follow-up weights

	Item number*	Estimate	SE	DEFF	DEFT
Proportions					
In vocation program	2	0.270	0.007	6.922	2.631
Worked last week	24	0.532	0.005	2.804	1.675
Working at clerical job	29	0.250	0.005	3.080	1.755
Current job is place where people goof off	33A	0.132	0.004	2.958	1.720
Work more enjoyable than school	33C	0.513	0.005	2.149	1.466
Job encourages good work habits	33D	0.789	0.004	2.114	1.454
Father non-professional	53A	0.887	0.005	6.276	2.506
Father finished college	55	0.213	0.007	7.040	2.653
Mother finished college	56	0.136	0.005	5.374	2.318
Watch more than one hour of TV per day	61	0.791	0.003	1.480	1.217
Career success important	73A	0.860	0.003	1.960	1.400
Having lots of money not important	73C	0.103	0.003	2.549	1.597
Important to be a leader in community	73F	0.476	0.006	3.748	1.936
Important to live close to parents	73H	0.707	0.005	3.147	1.774
Having leisure time not important	73L	0.017	0.001	1.552	1.246
Have a positive attitude toward self	75A	0.932	0.002	1.564	1.250
Good luck more important than hard work	75B	0.127	0.003	1.986	1.409
Believe someone or something prevents success	75E	0.256	0.005	3.122	1.767
Believe plans hardly ever work out	75F	0.199	0.004	2.434	1.560
Have little to be proud of	75L	0.126	0.003	1.992	1.411
Working to correct inequalities important	73J	0.396	0.004	1.738	1.318
No serious trouble with law	76A	0.949	0.003	4.845	2.201
Expect to finish full-time education	80	0.382	0.007	5.288	2.300
Would be satisfied with less than college ed.	82	0.744	0.006	4.693	2.166
Seen by others as physically unattractive	76	0.103	0.003	2.480	1.575
Married	97A	0.035	0.002	2.883	1.698
Expect first child by age 25	97B	0.538	0.005	2.404	1.550
Expect to have own home or apt. by age 24	97D	0.921	0.002	1.326	1.151
Expect to have no children	98	0.089	0.003	2.706	1.645
Hard of hearing	103C	0.019	0.001	1.472	1.213

Table 3.15--Standard errors and design effects associated with estimated proportions and averages of first follow-up sophomores who had specified characteristics, using first follow-up weights (continued)

	Item number*	Estimate	SE	DEFF	DEFT
Averages					
Vocabulary score		10.387	0.085	5.776	2.403
Reading score		7.657	0.072	5.217	2.284
Math, part 1 score		10.820	0.143	7.407	2.722
Math, part 2 score		2.736	0.041	5.031	2.243
Science score		9.475	0.073	5.969	2.443
Writing score		9.503	0.074	4.993	2.234
Civics score		5.441	0.037	4.326	2.080
MEAN (Proportion only)				3.136	1.719
MEAN (All statistics)				3.589	1.837
MINIMUM				1.326	1.151
MAXIMUM				7.407	2.722
STANDARD DEVIATION				1.804	0.470

* First follow-up questionnaire number.

Table 3.16--Standard errors and design effects associated with estimated proportions and averages of first follow-up sophomores who had specified characteristics using base year weights (BYWT)

	Item number*	Estimates	SE	DEFF	DEFT
		Proportions			
In vocational program	1	0.212	0.006	5.705	2.389
Worked last week	24	0.362	0.005	2.901	1.803
Working at clerical job	27	0.082	0.003	2.649	1.628
Current job is place where people goof off	30A	0.163	0.003	1.356	1.164
Work more enjoyable than school	30C	0.557	0.006	3.050	1.746
Job encourages good work habits	30D	0.722	0.003	0.945	0.972
Father non-professional	38	0.883	0.004	3.182	1.784
Father finished college	39	0.225	0.007	5.308	2.304
Mother finished college	42	0.139	0.005	4.508	2.123
Watch more than one hour of TV per day	48	0.909	0.003	2.896	1.702
Career success important	61A	0.850	0.003	1.846	1.359
Having lots of money not important	61C	0.102	0.003	2.556	1.599
Important to be a leader in community	61F	0.539	0.005	2.578	1.606
Important to live close to parents	61H	0.749	0.004	2.200	1.483
Having leisure time not important	73L	0.022	0.001	1.189	1.091
Having a positive attitude toward self	62A	0.909	0.002	1.131	1.064
Good luck more important than hard work	62B	0.155	0.003	1.612	1.270
Believe someone or something prevents success	62E	0.301	0.004	1.736	1.317
Believe plans hardly ever work out	62F	0.221	0.004	2.190	1.480
Having little to be proud of	62L	0.156	0.003	1.623	1.174
Working to correct inequalities important	61J	0.363	0.003	1.003	1.001
No serious trouble with law	67A	0.944	0.002	1.944	1.394
Expect to finish full-time education	69	0.397	0.006	3.916	1.979
Would be satisfied with less than college ed.	71	0.800	0.005	3.943	1.986
Seen by others as physically unattractive	67C	0.166	0.003	1.606	1.267
Married	78A	0.003	0.000	--	--
Expect first child by age 25	78B	0.583	0.004	1.563	1.250
Expect to have own home or apt. by age 24	78D	0.929	0.002	1.469	1.212
Expect to have no children	80	0.101	0.003	2.458	1.568
Hard of hearing	88C	0.024	0.001	1.034	1.017

Table 3.16--Standard errors and design effects associated with estimated proportions and averages of first follow-up sophomores who had specified characteristics using base year weights (BYWT) (continued)

	Item number*	Estimates	SE	DEFF	DEFT
		Averages			
Vocabulary score		8.479	0.068	4.070	2.017
Reading score		6.649	0.060	4.025	2.006
Math, part 1 score		9.801	0.116	5.646	2.376
Math, part 2 score		2.494	0.039	5.148	2.269
Score		8.777	0.069	5.540	2.354
Writing score		8.127	0.070	4.523	2.127
Civics score		4.479	0.039	5.182	2.276
MEAN (Proportion only)				2.417	1.508
MEAN (All statistics)				2.895	1.643
MINIMUM				0.945	0.972
MAXIMUM				5.705	2.389
STANDARD DEVIATION				1.523	0.448

*Base year questionnaire number.

3.8.2 Second Follow-Up

Table 3.17 displays the estimated percentages, standard errors, DEFFs, and DEFTs for variables from the second follow-up survey data. Since only ten of the thirty non-test items presented for the base year and first follow-up survey were included in the second follow-up survey questionnaire, twenty additional items, representing estimated proportions of varying magnitudes, were added to this table. Table 3.17 shows that the mean DEFT for the 30 estimated percentages from the second follow-up survey is 1.5, a smaller figure than observed for the first follow-up and about equal to that for the base year mean design effect calculated for proportions only (omitting test scores, which may be exceptionally influenced by the clustered sample design). The variability of the DEFFs across the thirty estimates is also much smaller for the second follow-up data than for prior waves, but this may be largely due to differences in the lists of items for which estimates, sampling errors, and design effects were calculated.

Table 3.17--Estimated percentages, standard errors and design effects in the percentages of the second follow-up sophomores who had specified characteristics (weight = FU2WT)

Statistic	Item number	Estimate	SE	DEFF	DEFT
Working full time, Feb '84	SY3A	58.51	0.67	2.53	1.59
Taking academic courses, Feb 84	SY3C	33.61	0.81	4.00	2.00
Looking for work, Feb 84	SY31	9.96	0.35	1.86	1.33
Currently married	SY56	12.31	0.47	2.77	1.66
Have one or more children	SY65A	11.80	0.43	2.18	1.48
Expect to have 3 or more children	SY64	33.92	0.55	1.78	1.33
Have served on military active duty	SY43	6.21	0.35	2.80	1.67
If in PSE 82-84:					
Earned no degree	SY181,J-20I,J	70.40	0.64	1.35	1.16
Earned vocational degree	SY181,J-201,J	1.11	0.14	1.23	1.11
Earned 4yr college degree	SY181,J-201,J	1.47	0.21	2.14	1.46
Enrolled in postsecondary education Oct 82	SEOC82	44.68	0.70	2.67	1.63
Enrolled in postsecondary education, Oct 83	PSEOC83	42.78	0.79	3.43	1.85
If employed: In clerical occupation, Oct 83	SY46A-49A	24.65	1.33	2.02	1.42
Employed, Oct 83	JOBSOC83	66.57	0.63	2.37	1.54
Have used pocket calculator	SY8A2-A4	90.71	0.39	2.42	1.56
Have used computer terminal	SY8B2-B4	47.49	0.74	2.77	1.66
Have used mainframe computer	SY8E2-E4	23.33	0.60	2.51	1.59
Have used video tape recorder	SY8F2-F4	53.82	0.59	1.76	1.33
Have used audio cassette deck	SY8H2-H4	88.26	0.40	1.97	1.40
Have used word processor	SY8I2-I4	9.09	0.40	2.56	1.60
Currently registered to vote	SY69	53.72	0.70	2.61	1.62
Have voted in election since turning 18	SY70	33.38	0.72	3.08	1.76
Being successful in job very important	SY71A	85.27	0.45	2.11	1.45
Marrying the right person very important	SY71B	87.63	0.41	2.03	1.43
Having lots of money very important	SY71C	29.40	0.64	2.61	1.61
Being a community leader very important	SY71F	10.04	0.40	2.34	1.53
Better opportunities for children very important	SY71G	72.66	0.56	2.05	1.43
Correcting inequalities very important	SY71J	14.08	0.50	2.78	1.67
Having children very important	SY71K	49.19	0.65	2.25	1.50
Having leisure time very important	SY71L	72.14	0.67	2.95	1.72
Mean				2.40	1.54
Minimum				1.23	1.11
Maximum				4.00	2.00
Standard Deviation				0.56	0.18

In general, the overall efficiency of the sophomore cohort second follow-up sample design appears to benefit from both a more proportionate allocation than in prior survey waves and from smaller cluster sizes. The second follow-up design decreased somewhat the disproportionality of the minority groups and other subsamples and decreased the relative variance of the sampling weights (RAWWT) from about 1.00 in the first follow-up to about 0.89 in the second follow-up. At the same time, the second follow-up design reduced the average cluster size from approximately 30 in the first follow-up to less than 15 in the second follow-up. Furthermore, the effects of the initial clusters on the efficiency of follow-up samples may be expected to diminish as sample members become more dispersed geographically and more differentiated in terms of life experiences.

The distributional statistics of the design effects and root design effects for the same 30 second follow-up items in Table 3.17 for the total population and 11 selected domains are shown in Table 3.6-5 of the Second Follow-Up Data File User's Manual. With the exception of Hispanics, the second follow-up DEFTs for subgroups were consistently smaller than for the total population. The relative efficiency of the Hispanic subsample continues to be differentially affected by the somewhat greater clustering of Hispanic sample members in specific schools within relatively few geographical areas. Moreover, the variability of the DEFTs for Hispanics was about twice that observed for most other subgroups. Thus, for analysis of data from Hispanics, the use of a single generalized design effect to inflate simple random sample estimates of sampling errors involves a greater amount of approximation.

3.8.3 Third Follow-Up

Standard errors, DEFFs, and DEFTs for 30 third follow-up survey items are shown in Table 3.18. The mean DEFT is 1.48, which is just slightly below the mean DEFT for the second follow-up. The variability of the DEFTs is lower for the third follow-up items (.10) than it was for the second follow-up items (.18). However, these statistics are not directly comparable because the method of calculating standard errors (and hence design effects) was different. In the second follow-up BRR estimates were employed while the third follow-up used Taylor series estimates.

Table 3.18--Estimated percentages, standard errors and design effects in the percentages of the third follow-up sophomores who had specified characteristics (weight = FU3WT)

	Item number	Estimate	SE	DEFF	DEFT
Working at full or part time job, Feb 86	TY3A	67.47	0.58	2.02	1.42
Taking academic courses, Feb 86	TY3C	26.84	0.63	2.68	1.64
Looking for work, Feb 86	TY3I	19.58	0.36	2.05	1.43
Currently married	TY41	23.14	0.56	2.36	1.54
Currently divorced	TY41	1.85	0.17	2.00	1.42
Currently have one or more children	TY49	22.33	0.58	2.55	1.60
Expect to have three or more children	TY48	31.72	0.60	2.16	1.47
In PSE 84-86: earned no degree	TY21I-22I	21.36	1.15	2.05	1.43
In PSE 84-86: received vocational degree	TY21H-22H	27.98	1.42	2.60	1.61
In PSE 84-86: received 4-year degree	TY21H-22H	31.36	1.35	2.22	1.49
Enrolled in PSE, Oct 84	TY21C-22C	32.11	0.66	2.64	1.63
Enrolled in PSE, Oct 85	TY21C-22C	28.36	0.61	2.45	1.56
In PSE 84-86: v. dissat. w/career counts	TY28E	5.52	0.41	2.07	1.44
In PSE 84-86: some sat. with curriculum	TY28I	50.41	0.84	1.78	1.33
Applied for grad/ professional school	TY39	4.46	0.28	2.23	1.49
If employed 84-86, 1st job clerical	TY8A	24.83	0.53	1.88	1.37
Had any job between 84-86	TY7	93.81	0.30	2.10	1.45
Did not receive unemployment-85	TY17D85	86.41	0.82	2.16	1.47
Currently registered to vote	TY56	66.40	0.67	2.58	1.60
Have voted since 1984	TY57	51.13	0.70	2.47	1.57
Active participant in service org.	TY59K	1.49	0.13	1.40	1.18
Job security very important	TY16C	75.74	0.56	2.13	1.44
Success in job very important	TY68A	79.88	0.51	2.03	1.43
Marrying the right person very important	TY68B	86.36	0.44	2.14	1.46
Having lots of money very important	TY68C	22.68	0.52	1.94	1.39
Being a community leader very important	TY68F	6.65	0.31	1.97	1.40
Providing better opp. for kids very imp.	TY68G	69.65	0.65	2.54	1.59
Correcting social inequalities very important	TY68J	11.02	0.42	2.32	1.52

Table 3.18--Estimated percentages, standard errors and design effects in the percentages of the third follow-up sophomores who had specified characteristics (weight = FU3WT) (continued)

	Item number	Estimate	SE	DEFF	DEFT
Having children very important	TY68K	47.85	0.64	2.08	1.44
Having leisure time very important	TY68L	68.21	0.59	2.05	1.43
Mean				2.19	1.48
Minimum				1.40	1.18
Maximum				2.68	1.64
Standard deviation				0.29	0.10

The distributional statistics of the design effects and root design effects for the same 30 second follow-up items in Table 3.18 for the total population and 11 selected domains are shown in Table 3.7-6 of the Third Follow-Up Data File User's Manual. The mean DEFFs and DEFTs for these domains are all very similar to those given below in Table 3.20 for the fourth follow-up.

3.8.4 Fourth Follow-Up

Standard errors, DEFFs, and DEFTs for 30 fourth follow-up survey items are shown in Table 3.19. The first 14 items also appear in Table 3.18. The mean DEFT for the fourth follow-up is 1.43, which is a little below the mean DEFT for the third follow-up (1.48). The fourth follow-up variability of the DEFTs (0.08%) is also a little below that of the third follow-up (0.10%).

Table 3.19--Estimated percentages, standard errors and design effects in the percentages of the fourth follow-up sophomores who had specified characteristics (weight = FU4WT)

	Item	Estimate	SE	DEFF	DEFT	N
Working at full- or part-time job?	2.1	79.27	0.50	1.92	1.39	12636
Now taking undergraduate courses?	2.5	5.42	0.29	2.07	1.44	12636
Currently looking for work?	2.2	3.85	0.26	2.31	1.52	12636
Married on 1/1/92?	8.1	51.62	0.66	2.17	1.47	12469
Divorced on 1/1/92?	8.3	6.64	0.31	1.93	1.39	12469
Have one or more children?	53	51.09	0.68	2.34	1.53	12640
Received 4-year degree since 1982?	32	22.40	0.61	2.70	1.64	12601
Applied to grad./prof. school(s)?	22	7.83	0.35	2.10	1.45	12383
Currently registered to vote?	62	64.97	0.67	2.47	1.57	12506
Voted in the last 12 months	61	33.54	0.60	2.03	1.43	12573
Active in a service organization?	59.9	4.01	0.23	1.74	1.32	12635
Success in job very important?	58.1	95.90	0.24	1.84	1.36	12526
Lots of money important?	58.2	55.57	0.60	1.82	1.35	12457
Better opport. for kids very impt.?	58.5	96.54	0.23	1.97	1.40	12435
Lives with spouse/partner?	4.1	56.86	0.65	2.17	1.47	12618
Now working on GED?	10.2	5.77	0.29	1.94	1.39	12564
Now taking postsecondary classes?	2.5-6	10.53	0.38	1.94	1.39	12636
Loans for education since HS?	28	26.72	0.56	2.00	1.42	12514
Highest education expected:						
Cert./lic./tech. award?	20.5	10.08	0.40	2.20	1.48	12434
Sales/marketing training since HS?	35.H	22.98	0.56	2.21	1.49	12463
Taken real estate licensing exam?	36.18	1.32	0.13	1.64	1.28	12640
Courses by mail/TV/radio/newspaper?	34.B	4.81	0.29	2.29	1.51	12489
Jobs are very diff. from training	45.A.3	38.58	0.61	1.94	1.39	12328
Employer-trained, last 12 months?	46	42.65	0.63	2.02	1.42	12438
Satisfied w/ job's pay/fringe?	52.A	70.39	0.61	2.20	1.48	12298
Satisfied w/ working conditions?	52.C	85.03	0.44	1.87	1.37	12301
Satisfied w/ job's supervisor?	52.F	83.21	0.48	2.00	1.41	12108
Supports person not immed. family?	56	6.13	0.28	1.71	1.31	12506
Has monthly mortgage payments?	66.1	40.98	0.61	1.91	1.38	12423
Has monthly auto loan payments?	66.2	48.50	0.68	2.30	1.52	12435
Mean				2.06	1.43	
Minimum				1.64	1.28	
Maximum				2.70	1.64	
Standard deviation				0.23	0.08	
Median				2.01	1.42	

Table 3.20 presents selected distributional statistics for the DEFFs and DEFTS for the same 30 fourth follow-up items contained in the table 3.19 for the total population and for 11 selected domains. For each of the 12 domains, the mean DEFFs and DEFTs are very close to the mean DEFFs and DEFTs for the same domain of the third follow-up.

Table 3.20--Distributional statistics for design effects and root design effects for 30 survey measures in 12 domains for the percentages of the fourth follow-up sophomores who had specified characteristics

		DEFF	DEFT
Total population	Mean	2.06	1.43
	Minimum	1.64	1.28
	Maximum	2.70	1.64
	Standard deviation	0.23	0.08
	Median	2.01	1.42
Hispanic	Mean	3.02	1.72
	Minimum	1.31	1.15
	Maximum	5.10	2.26
	Standard deviation	0.73	0.21
	Median	3.13	1.77
Black	Mean	2.25	1.50
	Minimum	1.50	1.22
	Maximum	3.44	1.85
	Standard deviation	0.39	0.13
	Median	2.23	1.49
Whites and others	Mean	1.81	1.34
	Minimum	1.45	1.21
	Maximum	2.43	1.56
	Standard deviation	0.21	0.08
	Median	1.84	1.36
Male	Mean	1.95	1.39
	Minimum	1.67	1.29
	Maximum	2.72	1.65
	Standard deviation	0.22	0.07
	Median	1.91	1.38
Female	Mean	1.97	1.40
	Minimum	1.72	1.31
	Maximum	2.26	1.50
	Standard deviation	0.15	0.05
	Median	1.97	1.40
Lowest quartile SES	Mean	1.93	1.39
	Minimum	1.39	1.18
	Maximum	2.47	1.57
	Standard deviation	0.24	0.09
	Median	1.97	1.40
Second quartile SES	Mean	1.82	1.34
	Minimum	1.33	1.15
	Maximum	2.86	1.69
	Standard deviation	0.32	0.11
	Median	1.79	1.34

Table 3.20--Distributional statistics for design effects and root design effects for 30 survey measures in 12 domains for the percentages of the fourth follow-up sophomores who had specified characteristics (continued)

		DEFF	DEFT
Third quartile SES	Mean	1.64	1.28
	Minimum	1.41	1.19
	Maximum	1.95	1.40
	Standard deviation	0.13	0.05
	Median	1.66	1.29
Highest quartile SES	Mean	1.79	1.33
	Minimum	1.46	1.21
	Maximum	2.43	1.56
	Standard deviation	0.23	0.08
	Median	1.71	1.31

The mean DEFTs for all the subgroups except Hispanics are no larger than 1.5. The mean estimated DEFT for Hispanics was 1.72, which is somewhat higher. The DEFTs for Hispanics continue to be affected by the somewhat greater clustering of the Hispanic sample members in specific schools and relatively few geographical areas. In addition, the variability of the DEFTs for the Hispanic sample across different items was also twice that observed for most of the other domains (standard deviation of .21 versus .10 or less). However, this variability by itself is not that great; the standard deviation of 0.21 is not much greater than the standard deviation exhibited by the DEFTs for all the domains combined in the second follow-up (0.18).

We also re-created Tables 3.19 and 3.20 using the panel weight instead of the fourth follow-up weight (the re-created tables are not included). Only those students who have been respondents for every survey wave have a non-zero panel weight. These tables were very similar to Tables 3.19 and 3.20. Because of the reduction in sample size when using the panel weight, both the mean of the actual standard errors (from 0.46 to 0.48), and the mean of the simple random sample standard errors (from 0.32 to 0.35) increased. Since the denominator (simple random sample standard error) increased slightly more, the DEFT ratio slightly decreased, from 1.43 to 1.37.

The preceding data and discussion lead to the conclusion that the analyst seeking an appropriate value to use for a root design effect to inflate simple random sampling-based estimates of sampling error may simply use 1.5. If the statistic is based largely on the Hispanic subsample, a root design effect of 1.75 will be more appropriate. If the statistic is more complex than a simple proportion or mean, the DEFTs just recommended will probably be conservative in that they will tend to overestimate the true standard errors.

3.8.5 Transcript Data Collection

We also chose 28 composite variables from the transcript study to examine design effects at the student level. The first seven of these variables are percentages, while the remaining 21 are continuous variables. Because students can have more than one transcript, the idea of a "completed case" is a little more complex. We examined these items under two different definitions of a "completed case." Separate weights were prepared under these definitions.

The first transcript weight, PSEWT1, was created for all students for which we received at least one transcript. PSEWT1 = 0 for all students with no transcripts received. Standard errors, DEFFs, and DEFTs for these 28 items are shown in Table 3.21 for all students with at least one transcript received. The mean DEFT of 1.40 is quite similar to that of the fourth follow-up questionnaire items (1.43). Table 3.21 shows that we received at least one transcript for about 8,400 students, or about two- thirds of fourth follow-up respondents.

Table 3.21--Estimated percentages, standard errors and design effects in the percentages and means of the sophomores with at least one transcript received who had specified characteristics (weight = PSEWT1)

	Estimate	SE	DEFF	DEFT	N
Two or more transcripts requested?	51.40	0.75	1.90	1.38	8447
No degree earned?	54.85	0.86	2.52	1.59	8447
Certificate is highest degree earned?	4.54	0.34	2.25	1.50	8447
AA is highest degree earned?	8.05	0.41	1.92	1.39	8447
BA earned?	32.57	0.83	2.65	1.63	8447
Business undergraduate primary major?	12.43	0.63	1.87	1.37	5127
Psychology undergraduate primary major?	2.81	0.29	1.58	1.26	5127
Time to BA in months?	53.79	0.49	1.89	1.37	1943
Time to AA in months?	36.93	1.32	1.86	1.36	481
Total # of undergraduate credits?	69.05	0.95	2.06	1.43	8239
Total undergraduate GPA?	2.72	0.01	2.06	1.44	7442
Total # of graduate credits?	4.33	0.25	1.77	1.33	8239
Total # of humanities credits?	11.06	0.21	2.14	1.46	8239
Total # of social science credits?	12.72	0.24	1.99	1.41	8239
Total # of science/engineering credits?	9.04	0.26	1.83	1.35	8239
Total # of business credits?	7.59	0.23	1.85	1.36	8239
Total # of personal development credits?	2.15	0.05	1.90	1.38	8239
Total # of all math credits?	5.18	0.10	1.83	1.35	8239
Total # of computer science credits?	3.09	0.12	1.92	1.39	8239
Total # of all foreign language credits?	1.84	0.08	1.97	1.40	8239
Humanities GPA	2.74	0.01	1.90	1.38	6376
Social science GPA	2.61	0.01	1.94	1.39	6150
Science/engineering GPA	2.50	0.02	1.91	1.38	4857
Business GPA	2.66	0.02	1.72	1.31	3530
Personal development GPA	3.28	0.02	1.97	1.40	4087
All math GPA	2.52	0.02	1.73	1.32	5308
Computer science GPA	2.77	0.02	1.72	1.31	3337
All foreign language GPA	2.93	0.03	2.09	1.44	2051
Mean			1.96	1.40	
Minimum			1.58	1.26	
Maximum			2.65	1.63	
Standard deviation			0.22	0.08	
Median			1.91	1.38	

Table 3.22--Distributional statistics for design effects and root design effects for 28 survey measures in 12 domains for the percentages and means of students in the sophomore cohort with at least one transcript received who had specified characteristics

		DEFF	DEFT
Total population	Mean	1.96	1.40
	Minimum	1.58	1.26
	Maximum	2.65	1.63
	Standard deviation	0.22	0.08
	Median	1.91	1.38
Hispanic	Mean	2.54	1.59
	Minimum	1.26	1.12
	Maximum	4.52	2.13
	Standard deviation	0.55	0.17
	Median	2.44	1.56
Black	Mean	2.01	1.41
	Minimum	0.88	0.94
	Maximum	2.88	1.70
	Standard deviation	0.38	0.14
	Median	2.03	1.43
Whites and others	Mean	1.75	1.32
	Minimum	1.43	1.20
	Maximum	2.28	1.51
	Standard deviation	0.18	0.07
	Median	1.72	1.31
Male	Mean	1.83	1.35
	Minimum	1.17	1.08
	Maximum	2.39	1.55
	Standard deviation	0.23	0.09
	Median	1.76	1.33
Female	Mean	1.84	1.36
	Minimum	1.60	1.27
	Maximum	2.27	1.51
	Standard deviation	0.15	0.05
	Median	1.83	1.35
Lowest quartile SES	Mean	1.91	1.38
	Minimum	1.19	1.09
	Maximum	2.80	1.67
	Standard deviation	0.38	0.14
	Median	1.92	1.38
Second quartile SES	Mean	1.66	1.29
	Minimum	1.19	1.09
	Maximum	2.24	1.50
	Standard deviation	0.21	0.08
	Median	1.61	1.27
Third quartile SES	Mean	1.62	1.27
	Minimum	1.37	1.17
	Maximum	2.11	1.45
	Standard deviation	0.15	0.06
	Median	1.61	1.27

Table 3.22--Distributional statistics for design effects and root design effects for 28 survey measures in 12 domains for the percentages and means of students in the sophomore cohort with at least one transcript received who had specified characteristics (continued)

		DEFF	DEFT
Highest quartile SES	Mean	1.82	1.35
	Minimum	1.54	1.24
	Maximum	2.35	1.53
	Standard deviation	0.19	0.07
	Median	1.80	1.34

For each of the 12 domains except the highest quartile in socio- economic status (SES), the mean DEFFs and DEFTs are slightly below the mean DEFFs and DEFTs for the same domains of the fourth follow-up questionnaire items. The mean DEFTs for all subgroups except Hispanics are no larger than 1.4. The mean DEFT for Hispanics is 1.59, substantially below the mean DEFT for Hispanics of the fourth follow-up questionnaire items. The variability of the Hispanic subgroup DEFTs is also smaller (.17 versus .21) and more similar to the other subgroups than it was for the fourth follow-up questionnaire items.

A second transcript weight, PSEWT2, was created for all students that responded to the fourth follow-up and for which we received all requested transcripts. PSEWT2 = 0 for all students for which we are missing at least one requested transcript (or if the student did not respond to the fourth follow-up). Standard errors, DEFFs, and DEFTs for the same 28 items as in the previous two tables are shown in Table 3.23. The mean DEFT of 1.37 is slightly below the mean DEFT using PSEWT1 (1.40). Table 3.11 shows that we received all requested transcripts for about 6,000 students, or about one-half of the number of fourth follow-up respondents.

Table 3.23--Estimated percentages, standard errors and design effects in the percentages and means of the sophomores with all requested transcripts received who had specified characteristics (weight = PSEWT2)

	Estimate	SE	DEFF	DEFT	N
At least two transcripts requested?	44.20	0.87	1.84	1.36	6004
No degree earned?	51.41	0.96	2.22	1.49	6004
Certificate is highest degree earned?	4.75	0.41	2.23	1.49	6004
AA is highest degree earned?	8.85	0.50	1.86	1.36	6004
BA earned?	34.98	0.92	2.23	1.49	6004
Business undergraduate primary major?	12.79	0.73	1.88	1.37	3934
Psychology undergraduate primary major?	2.66	0.33	1.65	1.29	3934
Time to BA in months?	54.14	0.55	1.83	1.35	1548
Time to AA in months?	36.37	1.47	1.87	1.37	378
Total # of undergraduate credits?	73.69	1.10	1.91	1.38	5834
Total undergraduate GPA?	2.74	0.01	1.91	1.38	5403
Total # of graduate credits?	4.73	0.29	1.54	1.24	5834
Total # of humanities credits?	11.53	0.25	2.05	1.43	5834
Total # of social science credits?	13.25	0.27	1.75	1.32	5834
Total # of science/engineering credits?	9.63	0.32	1.77	1.33	5834
Total # of business credits?	8.43	0.29	1.82	1.35	5834
Total # of personal development credits?	2.31	0.07	1.94	1.39	5834
Total # of all math credits?	5.54	0.12	1.71	1.31	5834
Total # of computer science credits?	3.37	0.14	1.83	1.35	5834
Total # of all foreign language credits?	1.85	0.09	1.87	1.37	5834
Humanities GPA	2.77	0.01	1.85	1.36	4629
Social science GPA	2.62	0.02	1.99	1.41	4485
Science/engineering GPA	2.51	0.02	1.78	1.33	3599
Business GPA	2.67	0.02	1.75	1.32	2681
Personal development GPA	3.30	0.02	1.91	1.38	3039
All math GPA	2.52	0.02	1.63	1.28	3930
Computer science GPA	2.80	0.02	1.66	1.29	2556
All foreign language GPA	2.95	0.03	2.18	1.48	1446
Mean			1.87	1.37	
Minimum			1.54	1.24	
Maximum			2.23	1.49	
Standard deviation			0.18	0.06	
Median			1.85	1.36	

Because of the decreased sample size, the mean of the design-corrected standard errors increases from .31 to .35. The mean of the simple random sample errors also increases, from .22 to .26. Since the denominator (simple random sample standard error) increased slightly more, the DEFT ratio slightly decreased, from 1.40 to 1.37.

Table 3.24 presents selected distributional statistics for the DEFFs and DEFTS for the same 28 transcript items contained in the preceding table for the total population of students with all requested transcripts received and for 11 selected domains. For all of the 12 domains, the mean DEFFs and DEFTs are equal to or less than the mean DEFFs and DEFTs for the same domain using PSEWT1 (table 3.22).

Table 3.24--Distributional statistics for design effects and root design effects for 28 survey measures in 12 domains for the percentages and means of students in the sophomore cohort with all requested transcripts received who had specified characteristics

		DEFF	DEFT
Total population	Mean	1.87	1.37
	Minimum	1.54	1.24
	Maximum	2.23	1.49
	Standard deviation	0.18	0.06
	Median	1.85	1.36
Hispanic	Mean	2.40	1.54
	Minimum	1.04	1.02
	Maximum	4.14	2.03
	Standard deviation	0.59	0.19
	Median	2.32	1.52
Black	Mean	1.96	1.39
	Minimum	0.90	0.95
	Maximum	2.70	1.64
	Standard deviation	0.40	0.15
	Median	2.00	1.41
Whites and others	Mean	1.71	1.31
	Minimum	1.45	1.20
	Maximum	2.08	1.44
	Standard deviation	0.16	0.06
	Median	1.70	1.30
Male	Mean	1.78	1.33
	Minimum	1.09	1.04
	Maximum	2.14	1.46
	Standard deviation	0.20	0.08
	Median	1.76	1.33
Female	Mean	1.79	1.34
	Minimum	1.48	1.22
	Maximum	2.03	1.43
	Standard deviation	0.14	0.05
	Median	1.80	1.34
Lowest quartile SES	Mean	1.92	1.38
	Minimum	1.01	1.01
	Maximum	2.69	1.64
	Standard deviation	0.35	0.13
	Median	1.93	1.39

Table 3.24--Distributional statistics for design effects and root design effects for 28 survey measures in 12 domains for the percentages and means of students in the sophomore cohort with all requested transcripts received who had specified characteristics (continued)

		DEFF	DEFT
Second quartile SES	Mean	1.62	1.27
	Minimum	0.86	0.93
	Maximum	2.12	1.46
	Standard deviation	0.22	0.09
	Median	1.65	1.28
Third quartile SES	Mean	1.61	1.27
	Minimum	1.23	1.11
	Maximum	1.97	1.40
	Standard deviation	0.15	0.06
	Median	1.63	1.28
Highest quartile SES	Mean	1.78	1.33
	Minimum	1.43	1.20
	Maximum	2.17	1.47
	Standard deviation	0.17	0.06
	Median	1.78	1.33

The proceeding data and discussion lead to the conclusion that the analyst seeking an appropriate value to use for a root design effect to inflate the simple random sampling-based estimates of sampling error for transcript items may simply use 1.4. However, if the statistic is based largely on the Hispanic subsample, a root design of 1.6 will be more appropriate. If the statistic is more complex than a simple proportion or mean, the DEFTs just recommended will probably be conservative in that they will tend to overestimate the true standard errors.

END NOTES

<1> For further details on the base year sample design see Frankel, M.; Kohnke, L.; Buonanno, D.; and Tourangeau, R. (1981), High School and Beyond Sample Design Report. Chicago: NORC.

<2> The sampling frame, defined as the universe of high schools in the United States, was obtained from the 1978 list of U.S. elementary and secondary schools of the Curriculum Information Center, a private firm. This was supplemented by NCES lists of public and private elementary and secondary schools. Information on racial composition was obtained from the 1976 and 1972 DHEW/Office of Civil Rights Secondary School Civil Rights Computer File of public schools and the National Catholic Education Association's list of Catholic schools. Any school listed in any of these files that contained a 10th grade, a 12th grade, or both was made part of the frame.

<3> Apart from substitution for schools that refused, there were a number of schools in the originally drawn sample that were "out-of-scope," failing to fit the criteria for inclusion in the sample. The sample was augmented through selection of an additional school for each out-of-scope school, within major strata. Most of the out-of-scope schools were area vocational schools, having no enrollment of their own, although they were listed in the frame as having enrollments.

<4> Tourangeau, R.; McWilliams, H.; Jones, C.; Frankel, M.; and O'Brien, F. (1983), High School and Beyond First Follow-Up (1982) Sample Design Report. Chicago: NORC.

<5> See Tables 2.4-1 through 2.4-4 of C. Jones and B. D. Spencer (1985), High School and Beyond Second Follow-Up (1984) Sample Design Report. Chicago: NORC.

<6> See Cochran, W. G. (1977), Sampling Techniques, Third Ed. New York: Wiley. p. 361.

<7> See p. A-4 of Tourangeau, R.; McWilliams, H.; Jones, C.; Frankel, M.; and O'Brien, F. (1983), High School and Beyond First Follow-Up (1982) Sample Design Report. Chicago: NORC.

<8> See Frankel et al. (1981), p. 93.

<9> See Frankel et al. (1981), p. 124.

<10> See Tourangeau et al. (1983), Chapter 4.

<11> See Tourangeau et al. (1983), Chapter 4, Tables 4.1 and 4.3.

<12> Kish, L. and Frankel, M. (1974), "Inference From Complex Samples,"Journal of the Royal Statistical Society: Series B (Methodological), Vol. 36, pp. 2-37.

4. DATA COLLECTION

4.1 Overview

To date, HS&B has compiled data from six primary sources: school administrators, teachers, students, parents of selected students, high school administrative records (transcripts), and postsecondary administrative records (transcripts and financial aid). In the 1980 base year survey, 1,015 secondary schools served as the primary sampling units for the study. The principal or headmaster of each school was asked to complete a school questionnaire and to provide materials essential for the sampling of students in the 10th and 12th grades.

In-school samples of approximately 36 students in each grade were asked to fill out a Student Identification Pages booklet (which included several items on the use of non-English languages as well as confidential identifying information) and a student questionnaire, and to take a timed cognitive (achievement) test. Teachers of selected students were asked to fill out brief Teacher Comment Forms containing 10 items on student traits and behavior.

During the fall following the base year survey, data were collected from over 7,100 parents of student respondents (about half of these were from each student cohort). These data focused primarily on parents' ability to finance postsecondary education for their children.

The first follow-up survey in the spring of 1982 added a second wave of data from 1980 seniors and sophomores. School administrators were again asked to complete a school questionnaire and to provide information on the secondary level course offerings and enrollments for their institutions. In the fall of 1982, high school transcripts were requested for a probability sample of approximately 18,500 members of the 1980 sophomore cohort. Both sophomore and senior cohort members were contacted for the second follow-up in 1984 and the third follow-up in 1986. In 1992, the fourth follow-up was conducted only with sophomore cohort members. Data and materials collected for all waves of HS&B are described below.

4.2 Data and Materials Collected from Schools and Teachers

School personnel supplied three broad types of information for HS&B: school questionnaires, course offerings and enrollments, and Teacher Comment Forms. School personnel were also asked to provide materials such as student rosters and class schedules, but these are not part of the public use data base and are not discussed here.

4.2.1 School Questionnaires

In both the base year and the first follow-up, principals and headmasters (or their designates) were requested to complete questionnaires asking for basic information on institutional characteristics such as type of control, ownership, total enrollment, proportions of students and faculty belonging to policy-relevant groups, participation in Federal programs, and per-pupil expenditures. This information is stored primarily in a separate data file that can be easily merged with student data files or the high school course offerings file described below. In addition, approximately 19 of the most basic school characteristics have been stored on the student data files in order to facilitate the classification of students according to their school environment.

School questionnaires were sought from all 1,015 participating schools during the base year survey. In the first follow-up survey, school data were requested from those schools still in existence as independent institutions (i.e., that had not closed or merged with other schools), and that still had members of the 1980 sophomore cohort enrolled. In a few instances, when students from a base year school were transferred en masse to a different school, or when two schools within a district merged, school questionnaires were sought from the schools then attended by the sampled students. In such cases, data from the new schools were stored on separate school records in the HS&B School Questionnaire data file, and were not physically merged with data for the original school. A link variable ("connecting school ID") is stored both in the record for each base year sample school that sent its students to a first follow-up "target school," and in the record for each "target school" indicating the ID of the base year school where the students were originally sampled and surveyed. Data from the new "target schools" can be merged easily with data records for the students who transferred in groups. No new school data were sought for students who transferred as individuals.

4.2.2 Teacher Comment Forms

Teacher Comment Forms were requested from all faculty members who had taught any HS&B sampled students during the 1979-80 academic year, but these data were collected only during the base year survey. Teacher Comment Forms asked for perceptions about whether each selected student would probably go to college, was working up to potential, seemed popular with others, had talked with the teacher about school work or plans, seemed to dislike school, had enough self-discipline to hold a job, or had a physical or emotional handicap that affected school work. Data from these forms have been compiled into separate files with over 19,000 forms for each of the two student cohorts.

4.2.3 Course Offerings and Enrollments: Academic Year 1981-82

During the first follow-up, school administrators were asked to provide materials that would allow the construction of a complete listing of all secondary level courses offered including enrollment figures for the 1981-82 academic year. This information was not requested in any prescribed format, and thus was received in a variety of forms. In many instances, schools were able to provide computer-generated printouts of Master Teaching Schedules. In others, it was necessary to merge information from several sources such as annotated course listings, catalogs, and enrollment records. Procedures were established to maximize the completeness and accuracy of these materials.

In the data file constructed from these documents, each school is represented by a block of records that indicate for each course offered a six-digit course identification number, the duration and timing of the course (e.g., year-long, first semester, third quarter), the credits earned for successful completion, and the total number of students enrolled in the course during the entire 1981-82 academic year. This data set can be merged easily with the School Questionnaire file, the Sophomore Data files, or the High School Transcripts (Sophomores) file. In both the Course Offerings and Enrollments and the High School Transcripts files, individual courses were coded using the Classification of Secondary School Courses (CSSC).

4.2.4 Data Collection Procedure: Schools and Teachers

In both the base year and first follow-up surveys, it was first necessary to secure a commitment to participate in the study from the administrator of each sampled school. In the case of public schools, the process was begun by contacting the chief state school officer (usually the state superintendent of schools) to explain both the objectives of the study and the data collection procedures (especially those for protecting individual and institutional confidentiality), and to identify the specific districts and schools selected for the survey. Once approval was gained at the state level, contact was made with District Superintendents and after district approval was granted, contact was then made with school principals. Wherever selected private schools were organized into an administrative hierarchy (e.g., Catholic school dioceses), approval was obtained at the superior level before approaching the school principal or headmaster.

Within each cooperating school, principals were asked to designate a School Coordinator who would serve as a liaison between the NORC HS&B staff and the school administrator and selected students. The School Coordinator (most often a senior guidance counselor) handled all requests for data and materials, as well as all logistical arrangements for student-level data collection on the school premises.

In the base year, the School Coordinator assisted in assembling the materials for student sample selection. In the first follow-up, the Coordinator reviewed the school sample and assisted in determining students' current enrollment status. Once the enrollment status was updated, the Coordinator assisted in locating current addresses for selected sophomore cohort school leavers (i.e., transfers, dropouts, and students who graduated ahead of schedule) and senior cohort base year survey nonrespondents.

School questionnaires were sent to coordinators in the fall of 1979 for the base year survey and in the fall of 1981 for the first follow-up survey of the sophomore cohort. Student survey sessions were conducted between February and June of 1980 for both the seniors and sophomores, and between February and June of 1982 for just the sophomores. In most cases, school questionnaires were completed and returned to NORC before the spring survey sessions. Most of the remainder were collected by NORC Survey Representatives who visited participating schools to conduct student survey activities. About one hundred additional school questionnaires were obtained in the fall of 1982, when schools were recontacted to supply student transcripts for a sample of 1980 sophomores. This additional contact with the schools also offered an opportunity to retrieve missing data from critical items in the school questionnaires.

In the base year, coordinators were asked to distribute some 67,000 Teacher Comment Forms to faculty members who might have taught HS&B sampled students during the 1979-80 academic year. Completed forms were returned by the teachers themselves in addressed, prepaid envelopes.

During the first follow-up survey of the sophomore cohort, coordinators were asked to assemble course offerings and enrollments data to be given to Survey Representatives at the time of the student survey sessions. Although nearly 90 percent of the schools provided course offerings information during the spring of 1982, the majority were not able to provide enrollment figures until the fall of that year, when the schools were recontacted for the sophomores' transcripts. Substantial numbers of schools could not provide enrollment data at all (see Table 4.1).

Finally, School Coordinators were notified during the first follow-up data collection period that they would be recontacted the following fall for their assistance in conducting the Student Transcript Survey for the sophomore cohort. Several months later, each coordinator was sent a packet of materials including a list of selected students and a reimbursement voucher to cover the costs of reproducing up to 36 (or 72 in the case of merged schools) high school transcripts for 1980 sophomores. (If selected students had transferred individually to schools not in the HS&B sample, transcript requests were sent directly to the principal of the last school the student was known to have attended.) Initial transcript requests were followed several weeks later by a combination of letters and telephone calls offering further assistance to each nonresponding school. Follow-up activities continued through January of 1983.

Table 4.1 displays the completion rates for school questionnaires (both waves), course offerings and enrollments data, and student transcript collection efforts. (Completion rates cannot be calculated for Teacher Comment Forms due to the absence of information on the total number of faculty members who had taught HS&B sampled students during the base year.)

Table 4.1--Response rates for school level data collection

	School questionnaires					
	Base year	First follow-up	Course offering data	Enrollment data	HS&B schools	Transfer schools
Number selected	1,015	992 (a)	992	992	992	890 (b)
Number responding	997	970	955	729	949 (c)	771 (d)
Response rate	98.2%	97.9%	96.3%	73.5%	95.7%	86.6%

a. Of the 992 schools from which full participation was sought in the first follow-up, 975 were among the initial 1,015 that participated in the base year, and 17 were included because they received en bloc transfers of all students from base year HS&B schools. Of the 975 base year schools eligible for the first follow-up, school questionnaires were obtained from 956 or 98 percent.

b. Transfer schools are defined as those to which 1,065 students had transferred as individuals.

c. Of the 949 schools that responded, 4 were unable to furnish transcripts because the sampled students had received a GED only and had not graduated.

d. Of the 771 schools that responded, 115 were unable to furnish transcripts because sampled students had never registered, transferred again, dropped out before earning credits, etc.

4.3 Student Data Collection

In the base year survey, a single data collection methodology - on-campus administration of questionnaires and tests to the entire sample of students from each selected school--was employed for both student cohorts. In the first follow-up, members of the younger cohort, nearly all of whom were then in the 12th grade, were resurveyed using methods similar to those of the base year survey. Members of the 1980 senior cohort were surveyed primarily by mail. Attempts were made to interview nonrespondents to the mail survey (approximately 25 percent) either in person or by telephone.

4.3.1 Base Year Data Collection

Base year student data were collected from students in 1,015 high schools between February 1 and May 15, 1980. Sophomores and seniors within each school were gathered in separate groups on scheduled survey days to complete the questionnaires and tests in one session. NORC Survey Representatives (often assisted by the School Coordinator) were present with each group to explain survey procedures and to answer questions.

An Orientation Day was held in each school, usually one to two weeks prior to Survey Day, in order to explain to sampled students the objectives of the study and to brief them on the voluntary nature of the study, the tasks involved in participation, and the procedures for protecting the confidentiality of their responses. Efforts were made during orientation sessions to identify all twins and triplets selected into the HS&B sample and to recruit the nonsampled twins and triplets into the study. Finally, a check was made during the orientation to see that parental permission forms had been obtained for all selected students in each school or district that required parental approval.

The first step for students in each survey session was to complete a Student Identification Pages (SIP) booklet, which requested information about how to locate the student if selected for future follow-up. To preserve student confidentiality, these documents were handled, shipped, and stored separately from all other student instruments. A section of the SIP booklet also contained several questions designed to identify students who had been exposed to languages other than English outside of formal school courses. Students having such exposure answered a special series of questions in the SIP about their use of and proficiency in the non-English language, as well as their bilingual education experiences. These data were processed into a separate file containing responses from over 11,300 students.

Students were then given one hour to complete the questionnaires. During this time, Survey Representatives scanned the completed SIP booklets for missing or incomplete responses. At the end of the allotted time, questionnaires were collected. Students were given a ten-minute break during which Survey Representatives reviewed the questionnaires for completeness. Further attempts were made to obtain any data missing from either the SIP booklets or the student questionnaires before students left the survey session.

The cognitive tests were administered following the completion of the questionnaires. Tests consisted of six timed segments. The Senior Test Booklet also included a series of items on student perceptions about the six subtests and how the student was feeling while taking the test.

After the testing, students with incomplete SIP booklets or questionnaires were asked to remain so that missing data could be captured. For certain questionnaire items considered crucial to the analytical objectives of the study, students were given the option of marking a special oval in the question field indicating that they did not wish to answer.

Following the survey session, NORC Survey Representatives made arrangements with School Coordinators to conduct make-up sessions for students who were unable to attend the first Survey Day. Survey Representatives then packaged all completed student questionnaires and test booklets for shipment to NORC's optical scanning subcontractor to convert student responses to machine- readable form. Student Identification Pages, parental permission forms (if necessary), and administrative documents were returned to NORC's central offices for processing and storage. Table 4.2 displays separately for each student cohort the numbers and percentages of students who completed base year questionnaires and tests.

Table 4.2--Base year data collection results by student cohort

	Number of selections	Completed questionnaire		Completed test	
		N	(%)	N	(%)
1980 Sophomores	35,723	30,030	(84.6)	27,569	(77.2)
1980 Seniors	34,981	28,240	(80.7)	25,069	(71.7)
Total	70,704	58,270		52,638	(74.4)

4.3.2 First Follow-up Data Collection: 1980 Sophomore Cohort

During the fall of 1981, School Coordinators reviewed printed rosters of HS&B sophomore cohort members originally selected at their schools and indicated which of the students were still enrolled at the same schools and which had transferred to another school, graduated early, or left school without graduating. School Coordinators were also asked to supply current name and address information for all individuals in the latter three categories, and then return the rosters to NORC. Students listed on the rosters had been previously annotated with a sampling flag or marker reflecting predetermined selection probabilities for several student strata. Individuals who were both flagged and identified by School Coordinators as dropouts, transfers, or early graduates were then confirmed as selections into the school leaver sample. School leavers who were not predesignated by sampling procedures were classified as ineligible for the first follow-up.

It is important to note that the first follow-up sample design specifications defines the eligibility of students for follow-up by their enrollment status as of the scheduled Survey Day at their base year schools. Thus, School Coordinators had to repeat the review of the original student rosters on Survey Day, and any changes in student status from the original roster review (e.g., students transferring or leaving school, dropouts returning to full-time school enrollment) were immediately implemented by Survey Representatives in accordance with sample design specifications. By the completion of the data collection period, 25,150 students had been classified as currently enrolled in base year schools (or designated receiving schools--see below), and 4,587 had been selected into the school leaver sample (1,290 transfers; 696 early graduates; 2,601 dropouts).

On-campus data collection arrangements were sought for all sophomore cohort members who were still enrolled in the schools they attended during the base year, or who had transferred as part of a class to another school in the same district. (This latter group included students who attended a junior high school during the base year, as well as those whose base year schools closed or merged with other schools not in the HS&B sample.) Survey Days were successfully arranged in 952 school buildings. However, a total of 40 schools declined to hold survey activities on-campus during regular school hours, but in most of these instances, administrators of noncooperating schools assisted the survey effort by reviewing student rosters, identifying school leavers, and updating address information for sophomore cohort members. Many officials assisted NORC Survey Representatives in securing alternative sites for survey sessions and in encouraging sampled students to participate in off-campus administrations.

Survey Days were conducted between February 15 and June 1, 1982, and activities generally paralleled those used in the base year. On the first scheduled survey day, teams of NORC Survey Representatives, assisted by School Coordinators, administered student questionnaires and tests to groups averaging 20 students in size. Make-up sessions were scheduled for all schools in which the student-level response rate was less than 95 percent. NORC Survey Representatives conducted about 60 percent of the make-up sessions while school coordinators conducted 40 percent. By the end of the data collection period, 96 percent of the students eligible for on-campus survey administration had been resurveyed.

Two alternative data collection strategies were implemented for students enrolled in the 40 schools that declined to allow on-campus sessions. Students enrolled in the 27 noncooperating schools located within 100 miles of at least one NORC Survey Representative were contacted by telephone, screened for current enrollment status, and, if not classified as a school leaver, invited to participate in a group survey session at a local public facility. The screening process also allowed Survey Representatives to confirm the status of school leavers who had been predesignated for follow-up and to invite them to survey sessions as well. Over 95 percent of the 719 students currently enrolled at these 27 refusing schools were resurveyed in this manner.

There was a final group of 13 nonparticipating schools located over 100 miles from NORC Survey Representatives, but administering the survey at these schools using similar methods would have required unjustifiably large expenditures. As appropriate, students in these schools were screened by telephone for their current enrollment status and recruited to participate. In these instances, however, eligible students were sent packets containing questionnaires, supplements, and other materials through the mail. A total of 340 students were found to be currently enrolled in these 13 schools, and about 89 percent returned completed questionnaires to NORC offices. Cognitive test data were not collected from these sophomore cohort members.

Off-campus survey sessions were held for 1980 sophomore cohort school leavers between February 20 and June 25, 1982. Because it was necessary to reconfirm the enrollment status of each student as of the first scheduled Survey Day at students' base year schools, off-campus group administrations were always scheduled after Survey Day at the schools where selected transfers, early graduates, and dropouts had formerly been enrolled. Once the respondents' enrollment and selection status was established, Survey Representatives contacted school leavers by telephone and invited them to take part in group sessions to be resurveyed and retested. All school leavers were offered monetary incentives for participation ($5 for filling out the follow-up questionnaire and $10 for taking the test), and were reimbursed (up to $10) for travel expenses to and from the survey sites. Off-campus survey administrations were conducted using procedures as similar as possible to those for on-campus sessions. Survey Representatives scan-edited completed questionnaires during the testing period and attempted to obtain missing or incomplete data before participants left the sites. Because the off-campus sessions typically involved only two to five school leavers, these administrations were handled by a single Survey Representative.

Although 85 percent of the participating school leavers were resurveyed in group administrations, a substantial minority could not attend scheduled sessions. Survey Representatives were able to personally interview and retest 465 of these individuals whose home addresses were close to areas where other survey activities were underway. In addition, 92 interviews were conducted over the telephone, and 60 completed questionnaires were returned by mail by school leavers whose residences were more than 50 miles from the closest Survey Representative. No first follow-up test data were obtained for the latter two groups. Table 4.3 displays data collection results separately for dropouts, transfers, and early graduates.

Table 4.3--First follow-up data collection results for sophomore cohort school leavers by student type

	Number of selections	Completed questionnaires		Completed tests	
		N	(%)	N	(%)
Dropouts	2,601	2,289	(88.0)	2,034	(78.2)
Transfers	1,290	1,170	(90.7)	1,073	(83.2)
Early graduates	696	643	(92.4)	595	(85.4)
Total	4,587	4,102	(89.4)	3,702	(80.7)

4.4 Collection of Student Transcripts

During the fall of 1982, high school transcripts were collected for a sample of 1980 sophomores. Approximately 18,500 students were selected using a disproportionate allocation that balanced the need to maximize the numbers of selections from policy-relevant subgroups (e.g., dropouts, racial and ethnic minorities, twins) against the need for statistical efficiency in the computation of nationwide estimates from the data. In the last week of September 1982, survey materials were sent to approximately 1,900 schools (including HS&B sample schools and schools to which 1980 sophomores had transferred). On November 4, 1982 follow-up calls to School Coordinators and principals were initiated and continued as necessary through January 1983.

Transcripts were received and processed for approximately 16,200 students (88 percent of the sample). The response rate for HS&B sample schools (92 percent) was significantly higher than that obtained for schools to which HS&B students had transferred (83 percent). Most often, transcripts were not obtained because of the absence of a signed form from a student authorizing school officials to release the transcript (affecting about 3 percent of the students), and district or school policy against releasing student transcripts for research purposes (affecting about 2 percent of students).

Student Transcript Data Files contain records for each student listing, for each secondary level course taken, a six-digit course identification number, the school year and term that the course was taken, the credits earned, and the final grade. Courses that are part of special curricula or programs (e.g., bilingual education, special education, programs for gifted students) are identified as such.

In addition, each student's record contains information on the student's rank in class, overall grade point average, numbers of days absent for each school year, number of suspensions, the date and reason the student left the school, and identifying codes and scores for any standardized tests taken by the student (SAT, PSAT, ACT, or Advanced Placement tests). This data file is not part of the student questionnaire and test score data file, but it can easily be merged with the latter by means of the common student identification number.

4.5 Second Follow-Up Data Collection: 1980 Sophomore Cohort

By the time of the second follow-up, the sophomore cohort was out of school and data were collected through a mailed questionnaire. To obtain correct addresses, an address update letter was mailed to members of both HS&B cohorts in November, 1983. The address update packet included a cover letter, address update form, return envelope, and newsletter. In December, trained telephone interviewers at NORC's central office began locating activities for the cases whose letters were returned as undeliverable. By the time the questionnaires were mailed, addresses had been found for all but about 300 members of both cohorts. These 300 cases were then sent to field interviewers for further locating attempts.

Second follow-up survey questionnaires were mailed to 14,825 members of the sophomore cohort in February, 1984. Along with the questionnaire, respondents received a cover letter, an instruction sheet, a place marker, a pencil, a response incentive check for $5, and an addressed, prepaid envelope for returning the questionnaire to NORC.

By the end of the third week after the mailing, 37.8 percent of the sophomores had returned their questionnaires. In order to obtain useful information on the effectiveness of thank-you and reminder postcards in boosting response rates, two different postcard mailings were scheduled. At the end of the third week, half the sample was sent a postcard, thanking them for sending in the questionnaire or encouraging them to do so. At the end of the seventh week, those respondents who had not yet mailed in their questionnaires received a telephone reminder followed by a postcard. Completion rates were compared at the end of week ten. Among the respondents who had been sent the reminder at the end of the third week, 56.9 percent had returned their questionnaires, while only 53.3 percent of the second group had returned their questionnaires. Hence, mailing the postcard at the end of the third week appeared to boost the response rate by about 4 percentage points.

By the beginning of the sixth week, 44.9 percent of the sophomore cohort had returned completed questionnaires. Compared to the first follow-up, many more sample members were declared temporarily unlocatable at this stage of data collection. They had either moved after the fall locating letter was sent out or had failed to report any change of address. Therefore, in order to trace nonrespondents, Survey Representatives had to spend considerable time obtaining additional locating information.

During week nine, telephone and personal interviews began. At this time, 9,043, or 60.6 percent, of the questionnaires had been received. Telephone and personal interviews continued into August 1984, at which time the field period was closed. The final number of completed questionnaires for the sophomore cohort was 13,682, or 92 percent of the sample of 14,825. About 79 percent of the respondents completed and sent in questionnaires without assistance (self-administered); 15.6 percent were interviewed by telephone; and 5.3 percent were interviewed in person. Tables 4.4 and 4.5 display second follow-up data collection results by student type and sampling strata.

Table 4.4 -- Second follow-up data collection results by student type, sophomore cohort

Student response type	Initial selections	Completed cases	Refusals	Other*	Resp. rate
Stayed in HS	11,013	10,341	181	491	93.9%
Dropouts	2,584	2,219	60	305	85.9%
Transfers	752	679	15	58	90.3%
Early graduates	476	443	7	26	93.1%
Total	14,825	13,682	263	880	92.3%

*Included under "other" are cases that were not available, not located, deceased, or genuine other.

Table 4.5--Second follow-up data collection results by sampling strata, sophomore cohort

Sampling stratum	Initial selections	Completed cases	Refusals	Other*	Response rate
Cuban/Puerto Rican	990	890	18	82	89.9%
Hispanics - high achievement	886	844	13	29	95.3%
Hispanics - other	1,375	1,247	28	100	90.7%
Blacks - high achievement	741	688	10	43	92.8%
Blacks - other	1,295	1,176	16	103	90.8%
Asian/Pacific Islander	430	394	6	30	91.6%
American Indian/ Alaskan	292	260	2	30	89.0%
White - low SES/ high achievement	388	362	8	18	93.3%
White - other	8,428	7,821	162	445	92.8%
Total	14,825	13,682	263	880	92.3%

*Included under "other" are cases that were not available, not located, deceased, or genuine other.

4.6 Third Follow-Up Data Collection: 1986

In October 1985, NORC mailed a locating packet to members of the HS&B sample, excluding the deceased, the mentally incapacitated, and participants who had refused participation or could not be located during the second follow-up survey. The packet included a report about previous surveys, a letter of introduction, and an address form with space to update address information. NORC received a total of 10,346 (40 percent) responses to the mailing, with 6,593 updated addresses and 3,753 address verifications, and these were used to make corrections on the name and address file.

Locating packets that were returned as undeliverable were routed to an in-house telephone locating shop. Of 1,925 undeliverables, telephone interviewers were able to find addresses for 1,454, or 70 percent. The remainder were eventually sent to the field staff for more intensive locating.

Cases that had been declared unlocatable (1,017) during the second follow-up were sent directly to the field staff for locating. Of the 1,488 cases assigned to the field staff (these 1,017 plus the 471 for whom addresses could not be obtained by telephone), updated addresses were obtained for 418 (28 percent) respondents. These addresses, as well as forwarding addresses from the post office, were also entered on the name and address file.

Data collection began in the last week of February 1986 and continued through mid-September. For the first time, sophomores and seniors received the same questionnaire and for administrative purposes could be treated identically. Questionnaire packages were mailed to 26,820 respondents whose addresses had been updated during the prefield locating period. Packages contained questionnaires, a cover letter, a $5 respondent check fee, a pencil, and a return envelope. Survey materials were mailed first class with "Address Correction Requested" specified on envelopes.

By the end of the third week, 37 percent of the total sample had completed and returned their questionnaires. Those respondents who had not returned their questionnaires by the third week were sent follow-up postcards to thank those who had completed and returned their questionnaires and to encourage the others to send them in promptly. Because of the good effects evidenced during the second follow-up experiment, this card was sent to all respondents.

Telephone prompting of those who had not sent in questionnaires began in early April, approximately two weeks after postcards were mailed. NORC field interviewers contacted respondents to urge them to complete and return questionnaires. Offers to remail survey materials were made to those who reported they had not received questionnaires or had misplaced them.

While the field staff continued to contact respondents and encourage the self-administration of questionnaires, administration by telephone and in person began in June, during week 14 of the field period. At this time, 16,270, or 60.7 percent, of the questionnaires had been received. The number of cases completed with interviewer assistance began to increase in July and soon became the dominant method of administration. This continued through mid-September.

After 27 weeks, data collection ended with a final completion rate of 89.5 percent, or 23,993 completed questionnaires. The final completion rate for sophomores was 90.6 percent, or 13,429 completed questionnaires. The final completion rate for seniors was 88 percent, or 10,564 completed questionnaires. Table 4.6 displays the final completion rates for the sophomore sample by sampling strata.

Table 4.6--Data collection for the sophomore cohort by sampling, strata, third follow-up

	Initial selections	Completed cases	Refusals	Other	Response rate (%)
Cuban/Puerto Rican	990	829	20	141	83.7
Hispanic-high achievement	886	843	11	32	95.1
Hispanic-other	1,375	1,223	33	119	88.9
Black-high achievement	741	660	20	61	89.1
Black-others	1,295	1,123	25	147	86.7
Asian/Pacific Islander	430	385	6	39	89.5
American Indian/ Alaskan	292	252	7	33	86.3
White-low SES/ high achievement	388	360	6	22	92.8
White-others	8,428	7,750	185	493	92.0
Total	14,825	13,425	313	1,087	90.6

* Included under "other" are cases that were not available, not located, or deceased.

4.7.1 Fourth Follow-Up Data Collection: 1980 Sophomore Cohort

The fifth round of data collection for HS&B marked a change in data collection procedures. For the first time, a Computer Assisted Telephone Interview (CATI) was used in collecting data on the 1980 Sophomore Cohort. The CATI program used contained two instruments: the first instrument was used to locate and verify the identity of the respondent, while the second instrument contained all of the survey questions. The two instruments were linked so that with a few key strokes, an interviewer could move easily between them. This arrangement maximized system performance by not requiring the interviewer to access the large survey instrument until the respondent was on the telephone and had agreed to proceed with the interview.

Final testing of the CATI instrument with pretest respondents was completed by January 26, 1992. Because minor problems were detected by the interviewers; final programming only entailed transforming the introductory module into conversational interviewing. On February 5, letters were sent to all respondents with known telephone numbers to inform them that in the coming weeks they would be contacted to complete an interview for the HS&B fourth follow-up. Another set of letters were sent to respondents without telephone numbers requesting that they contact NORC on its toll free number. By February 14, data collection had begun.

The average administration time for an interview was 30.6 minutes. Some adjustments were made to the instrument in the interest of clarity and efficiency in interviewing. No further modifications were made to the CATI screens beyond May. By April, it was apparent that there were complications in locating sample members for interviews. Only 40 percent of the interviews had been completed, which did not meet the anticipated 50 percent targeted to be completed after 10 weeks. These difficulties had implications for both the schedule and the costs for the data collection task. They limited operations and necessitated extensive locating procedures, including a field staff to work cases that could not be completed in the telephone center.

In order to estimate the extent of the locating problems. A random subsample of cases was selected for tracking. The information obtained from this test was used to refine locating procedures and methods used by the telephone center and in the field.

Specialized training of interviewing staff in locating procedures was also undertaken. On April 29th, the initial locator training was conducted with five interviewers. Interviewers were introduced to four electronic resources: CBI, TRW, Compuserve and Trans Union. The interviewers were also given detailed information about the other resources used to locate respondents. The staff of locators grew to 43 persons by August.

Intensive field intervention was another method employed to locate respondents. At the time the field was brought on, the completion rate was 70.5 percent. The field staff used its resources to locate respondents and urged them to contact the central office to complete an interview. Overall, the field effort resulted in the location of 2140 sample members. The combined phone center and field locating efforts resulted in an overall completion rate of 85.3 percent.

Table 4.7--Data collection for the sophomore cohort by sampling strata, fourth follow-up

	Initial selections	Completed cases	Refusals	Other	Response rate
Cuban/Puerto Rican	990	764	32	194	77.2
Hispanic-high achievement	886	806	23	57	91.0
Hispanic-other	1,375	1,111	37	227	80.8
Black-high achievement	741	612	10	119	82.6
Black-others	1,295	982	23	290	75.8
Asian/Pacific Islander	430	356	9	65	82.8
American Indian/ Alaskan	292	239	4	49	81.8
White-low SES/ high achievement	388	356	9	23	91.8
White-others	8,428	7,414	235	779	88.0
Total	14,825	12,640	382	1,803	85.3

* Included under "other" are cases that were not available, not located, or deceased.

5. DATA CONTROL, PREPARATION AND PROCESSING

Data control and preparation refers to a series of procedures governing the conversion of completed questionnaire data to machine-readable form. The process involves monitoring the receipt of completed documents from respondents and the field interviewing staff; editing completed instruments for missing information and proper adherence to routing or skip instructions; assigning numeric codes to responses such as institutions attended, occupations, military specialties, and so on; retrieving missing information and resolving inconsistencies in responses to specified questions; and validating a percentage of the interviews conducted in person or by telephone.

5.1 Base Year Procedures

The base year procedures for data control and preparation differed significantly from the first and second follow-ups. Since the base year student instruments were less complex (for example, they employed only one skip pattern in the senior questionnaire and required no open-ended coding), the completed documents were sent by NORC Survey Representatives directly from the schools to the scanning subcontractor. The scanning computer was programmed to perform the critical item edit (described below) and to produce reports that identified the critical items with missing information for each case. The reports were sent to NORC where data retrieval was completed. (The Base Year Teacher Comment Forms were also sent directly to optical scanning, but no data retrieval was conducted.)

The base year school questionnaires and base year parent questionnaires were converted to machine readable form by the conventional key-to-disk method at NORC. In the base year, most school questionnaires were completed and returned to NORC before the scheduled Survey Day at the school; the remainder were collected by Survey Representatives during their Survey Day visits. This sequence permitted collection of missing school questionnaire data for most institutions during the course of scheduled survey activities, obviating the need for additional contact with school officials.

5.2 First Follow-Up Procedures

5.2.1 Shipping and Receiving Documents

Documents shipped from the field to NORC were assigned disposition codes that characterized the completion status for each case in terms of both respondent type and the presence or absence of relevant materials. Any discrepancies were resolved with the appropriate Survey Representatives or Field Managers within a 24-hour period. Data control disposition codes were then entered into the in-house processing segment of NORC's Automated Survey System (NASS), and cases were routed to the appropriate processing station. Additional updates were made to the NASS record for each case as the remaining procedures (editing and coding, data retrieval, interview validation) were completed.

5.2.2 Editing and Coding

A staff of 16 coder/editors handled over 40,000 student questionnaires (both cohorts) and nearly 1,000 school questionnaires. Editors and coders were trained for one week, and formal training was followed by a 100 percent review of the first 40 cases edited by each trainee. Those not performing satisfactorily were either terminated or retrained, depending upon the severity of the problem.

The first data preparation step for each completed document was the critical item edit. (The large sample and lengthy data collection instruments of HS&B made 100 percent editing of each questionnaire infeasible.) Approximately 40 items in each of the major survey instruments were designated as "critical, or "key" items. Items were so designated if they were deemed to be crucial to the methodological or analytical objectives of the study. Most of the key items are of self-evident policy relevance; others were chosen as a means of checking whether survey respondents had properly followed routing instructions, or whether they had inadvertently skipped portions of the questionnaires. Cases were deemed to have failed the critical item edit if the respondent did not provide a codeable response to any single key item. Thus, omissions, illegal multiple responses, and vague, unclear responses were grounds for failure. Items failing the edit were flagged and routed to the data retrieval station. There, respondents were called to obtain missing information or otherwise resolve the edit failure. In addition to the critical item edit, the following coding tasks were performed:

1. Occupation and industry were coded according to the US Department of Commerce, Bureau of the Census Classified Index of Industries and Occupations 1970, and the US Department of Commerce, Bureau of the Census Alphabetical Index of Industries and Occupations 1970. The 1970 edition was used so that the coding in HS&B would coincide with that used on NLS-72.

2. Postsecondary schools were coded using six-digit PSVD and FICE codes. The directories included the NCES Directory of Postsecondary Schools with Occupational Programs, 1975-76 and the NCES Education Directory, Colleges and Universities, 1981-82. Codes were created for unique schools not listed in these directories.

3. Military codes for specialized schools, specialty, and pay grade were classified according to the Department of Defense Occupational Conversion Table, Enlisted 1974, so that HS&B military coding would be compatible with that used for NLS-72.

4. The major field of study indicated for each postsecondary school attended was converted to a six digit code using the HEGIS Taxonomy. The directories used included: HEW, NCES, A Taxonomy of Instructional Programs in Higher Education, 1970, NCES, Standard Terminology for Curriculum and Instruction in Local and State School Systems (Handbook VI); and HEW, Vocational Education and Occupations, 1969. These directories were also used for field of study coding on NLS-72.

5. Open-ended questions in the Early Graduate and Transfer Supplements were coded.

6. To ensure compatibility with NLS-72, the various licenses, certificates, and other diplomas received by respondents were coded according to two-digit values created for the earlier study.

7. Numerical responses were transformed to darkened ovals to facilitate optical scanning.

5.2.3 Data Retrieval and Validation

The proportion of cases requiring retrieval varied widely between the sophomore and senior cohorts because of differences in the method of administration. Senior instruments completed in an unsupervised setting had a 43.6 percent retrieval rate, while sophomore instruments were notably below that at 16.5 percent. The lower retrieval rate among sophomores was achieved through the use of on-site edits performed by Survey Representatives on school Survey Days and at off-campus group administrations. Questionnaires with missing or incomplete information on critical items were handed back to the students for correction, and the students generally complied, time and circumstances permitting.

Interview validation requires the recontacting of selected respondents in order to repeat the collection of specified data. Data from validation calls (conducted from the central office) are then compared with data collected by Survey Representatives through personal or telephone interviews. Discrepancies in the two data sources were investigated, and if they could not be resolved, the respondent was reinterviewed. Additional cases were validated for an interviewer whenever a single validation failure occurred and follow-up action was taken as appropriate.

Since the process of validating an interview requires a phone call to the respondent, cases requiring both validation and retrieval were handled in a single call to lessen respondent burden. As noted earlier, approximately 10 percent of the instruments completed in person or by telephone were validated. In the first follow-up, no cases were found to fail validation checks.

5.3 Second Follow-Up Procedures

5.3.1 Shipping and Receiving Documents

Respondents and field interviewers mailed questionnaires to NORC's central office in Chicago. Arriving documents were sorted according to disposition codes that identified completed cases by method of administration (i.e., self-administered, telephone interview, or personal interview), and these disposition codes were then entered into NORC's Automated Survey System (NASS). As cases were routed through the data preparation system, an additional in-house update was made to the NASS record as each procedure (editing, coding, and retrieval) was completed. Codes designating validation cases were also entered. A final entry into the NASS record was made when the cases were processed for shipment to the scanning contractor. As in the first follow-up survey, a detailed transmittal listing every case in each carton accompanied every shipment to the optical scanning firm.

5.3.2 Editing and Coding

A staff of 12 coder/editors processed nearly 26,000 student questionnaires (both cohorts). Coder/editors were trained for 1-1/2 days. After a 100 percent review of the first 20 cases, coders not meeting quality control standards were either reassigned or retrained.

As in the first follow-up, the first data preparation step to be completed was the critical item edit. A list of 37 items in each survey instrument were designated as "critical" or "key" items.

The second follow-up survey marked the first time that respondents entered and filled in optically scanable grids for all of their answers to numeric questions. Therefore, in addition to the critical item edit, all numerical responses were examined for correct entry (e.g., right justification, omission of decimal points).

Other data preparation tasks included coding occupational and industrial information and licenses and certificates. Military specialized schooling, specialty, and pay grade were coded using a Department of Defense (DOD) coding scheme, Occupational Conversion Table, Enlisted Officer Civilian (December 1982). However, each DOD Officer Code, a numerical value followed by an alphabetical value, had to be converted to a three-digit number. To ensure that officer codes are not mistaken for enlisted codes, a "flag" was been placed at the beginning of each respondent file where an officer code was present. Coast Guard training and assignments received appropriate Navy codes, a procedure used by the Defense Manpower Center.

Coding of the names of postsecondary schools attended by respondents was accomplished by using the NCES Education Directory, Colleges and Universities, 1982-1983 and an updated source for vocational school programs, the NCES Directory of Postsecondary Schools with occupational Programs, 1982. As in the first follow-up, codes were created for schools not listed in these directories. The field of study information was coded using A Classification of Instructional Programs (CIP). Produced by NCES in 1981, this directory replaced A Taxonomy of Instructional Programs in Higher Education (HEGIS Taxonomy, 1972) and the Standard Terminology for Curriculum and Instruction in Local and State School Systems (known as Handbook VI), which were used in the first follow-up. To provide continuity between the first and second follow-ups, crosswalks between the HEGIS Taxonomy and the CIP, and between Handbook VI and the CIP were created.

5.3.3 Data Retrieval and Validation

The proportion of cases requiring missing data retrieval or other fail-edit callbacks for each cohort was similar: 29.1 percent for the older cohort and 32.5 percent for the younger cohort. Though it appears that second follow-up sophomore retrieval rates rose dramatically from the first follow-up, the comparison is misleading. First follow-up questionnaires received an on-site edit by Survey Representatives, and questionnaires with missing or incomplete information were returned to the respondents for completion. No on-site edit was possible during the second follow-up survey.

During the second follow-up, field supervisors conducted validation interviews with 10 percent of the respondents who had been interviewed on the telephone or in person. The data collected were then compared to questionnaire data. As in the first follow-up, no cases failed validation checks. (For a description of validation procedures, see Section 5.2.3.)

5.4 Third Follow-Up Procedures

5.4.1 Shipping and Receiving Documents

Respondents and field interviewers mailed questionnaires to NORC's central office in Chicago. Arriving documents were sorted according to disposition codes that identified completed cases by method of administration (i.e., self-administered, telephone interview, or personal interview). These disposition codes were then entered into NORC's Survey Management System (SMS) a microcomputer-based system that replaced the NORC Automated Survey System (NASS) used on earlier rounds of the study.

At the time of entry, the SMS generated and automatically entered the date each case was received. As cases were routed through the data preparation system, an additional in-house update was automatically made to the SMS record file as each editing, coding, and retrieval procedure was completed. A final entry into the SMS record was made when the cases were ready to be processed for shipment to the scanning contractor, Questar Data Systems.

5.4.2 Coding and Computer Assisted Data Entry

Coders were trained for two days, after which 100 percent of their first 20 cases were reviewed. If a coder's work did not prove to be satisfactory during this review, he or she was reassigned or retrained. A staff of four coders processed 23,993 student questionnaires (from both cohorts).

For this follow-up, coders were not responsible for editing responses; all editing was done using NORC'S Computer Assisted Data Entry (CADE, see below). Coders assigned values to the open-ended questions concerning occupation, industry, postsecondary school, and field of study. Occupation and industry codes were obtained from the U.S. Department of Commerce, Bureau of the Census's Classified Index of Industries and Occupations, 1970; and Alphabetical Index of Industries and Occupations, 1970, the same sources that were used in the previous follow-ups. Coding the names of the postsecondary schools attended by the respondents was accomplished using the HEGIS and Postsecondary Career School Survey Files provided by NCES. This file is the result of merging HEGIS codes from the NCES Education Directory, Colleges and Universities, published in the years 1981-1982 through 1985-1986, and the NCES Directory of Postsecondary Schools with Occupational Program, 1979 and 1981. As in the preceding follow-ups, codes were created for schools that did not appear in these directories. Codes beginning with 800000 were assigned to unlisted foreign schools, and codes beginning with 850000 were assigned to unlisted business and trade schools. Field-of-study information was coded using A Classification of Instructional Programs (CIP), as in the second follow-up.

In the third follow-up, for the first time, all codes were loaded into a computer program for more efficient access. Coders typed in a given response, and the program displayed the corresponding numerical code. This computerized coding system proved to be much faster and more accurate than manual look-ups.

The third follow-up survey marked the first time in the history of HS&B that numeric and critical items were key entered by individual operators rather than being scanned. Using a CADE program, operators were able to combine data entry with the traditional editing procedures. The CADE system, an offshoot of CATI (Computer Assisted Telephone Interviewing), steps question-by-question through critical and numeric items, skipping over questions that were slated for scanning and questions that were legitimately skipped because of a response to a filter question. Ranges were set for each question, preventing the accidental entry of illegitimate responses.

The CADE program accepted reserved codes to indicate a missing or illegitimate response. These codes were then converted to the standard reserved codes used in previous waves. To lessen the possibility of error, the CADE program required double entry of reserved codes on all critical questions.

Twelve CADE operators were trained for two days. After a 100 percent check of the first 20 cases, operators not meeting quality control standards were either terminated or retrained, depending on the severity of the problem. After the initial training period, a high percentage of cases continued to be checked until each operator met the appropriate standards.

CADE operators were responsible for the critical item edit, and those critical items that did not pass the edit were flagged for retrieval, both manually and by the CADE system. Numeric items, open-ended items, and filter items were also designated for CADE entry. These items have traditionally caused difficulty for respondents, particularly difficult have been numeric items, because respondents frequently have not right justified values or filled in grids correctly. Because these items were directly entered by operators who were inspecting each questionnaire, respondent errors could be discovered and resolved on an individual basis rather than through the more aggregate procedures of machine editing. After a missing critical item was retrieved by telephone interviewers, the questionnaire was returned to CADE for entry of the retrieved data. After completing "RE-CADE," questionnaires were checked and boxed for shipment to the scanning firm.

5.4.3 Data Retrieval and Validation

Critical-item retrieval was done by central office telephone interviewers through September 1986. With the retrieval rate at 41 percent, interviewers processed 7,167 questionnaires, retrieving items on 5,901 (86 percent). Of the remaining cases, 154 persons refused to answer the critical item(s) and 806 persons were considered unlocatable. A postcard listing a toll-free number was sent to the last known-address of unlocatable respondents; respondents called the toll free number in response to that mailing.

Validation procedures for the third follow-up centered on verifying data quality through item checks and verifying the method of administration for 10 percent of each interviewer's work. Each field manager was assigned a random number between 0 and 9 and validated each nth case for all her interviewers. Field managers telephoned the respondent to check several items of fact and to confirm that the interviewers had conducted a personal or telephone interview, or had picked up a questionnaire as indicated in the interviewer's report. Cases administered by Valdes Research, an independent contractor for Hispanic interviews, were validated from the central office at a rate of 30 percent. No cases failed validation.

5.5 Fourth Follow-Up Data Control and Processing

5.5.1 Computer-Assisted Telephone Interviewing (CATI)

The AutoQuest CATI system presented the instrument questions on a series of screens, each with one or more questions. Between screens, the system examined the responses to completed questions and used that information to route the interviewer to the next appropriate question. It also applied the customary edits--valid ranges, data field size and data type (e.g., numeric or text), and consistency with other answers or data from previous rounds. If it detected an inconsistency because of an interviewer miskey, or if the respondent simply realized that he or she made a reporting error earlier in the interview, the interviewer could go back and change the earlier response. As the new response was entered, all of the edit checks that were performed at the first response were again performed. The system then worked its way forward through the questionnaire using the new value in all skip instructions, consistency checks, and the like until it reached the first unanswered question, and control was then returned to the interviewer. In addition, when problems were encountered, the system could suggest prompts for the interviewer to use in eliciting a better or more complete answer.

Interviewers also received some additional coding capabilities by temporarily exiting the CATI program and executing separate coding programs (See Sections 5.5.4 and 5.5.5). Interviewers had programs to assist them in coding the respondents' postsecondary educational institutions, their occupations, and industries in which they were employed. Data from the coding programs were automatically sent to the CATI program for inclusion in the dataset.

At the conclusion of an interview, the completed case was deposited in the database ready for analysis. There was minimal post data entry cleaning for these data because the interviewing module itself conducted the majority of necessary edit checking and conversion functions.

5.5.2 Case Delivery to Interviewers

The main survey employed two modes of case delivery. The first method was controlled and monitored by the Telephone Number Management System (TNMS), a component of the CATI system. In the second method, TNMS record data for each noncomplete case was printed and case folders were created for hard copy case management.

5.5.3 Telephone Number Management System

The TNMS delivered cases to interviewers and controlled the flow of cases through the Telephone Center. For example, once a respondent had been contacted, the TNMS automatically placed the interviewer into the CATI interviewing module. If a respondent stopped an interview midstream, the data collected to that point was stored, and when the respondent was next contacted the case was presented to the interviewer at the breakoff point. The TNMS automatically delivered cases to interviewers based on prior appointments, interviewer availability, and the result of past attempts. Telephone numbers were delivered based on a set of scheduling rules that were customized for the demographics of the HS&B sample. For example, initial calls to residential numbers were scheduled for delivery to CATI operators in the evening to maximize the probability of contacting the respondent and to take advantage of lower telephone rates. The scheduler then routed active telephone numbers through different time periods in order to maximize the chance for contacting the respondent. Cases were staggered based on the respondent's time-zone so that most attempts were made during peak contacting times.

There were 149 preloaded data items for each TNMS case record; on other surveys, a TNMS record has contained anywhere from 10 to 15 preloaded variables. The information contained in each TNMS case record was vital to conducting the HS&B interview. However, the relatively large record size increased the amount of time it took TNMS to process all transaction types; supervisor functions, like reporting, case review and modification, were affected along with case delivery and routing.

Initially, TNMS was organized to allow cases to be worked in distinct phases:

1. calling respondent numbers;

2. calling Directory Assistance for a number for the respondent at his/her last known address, from the Third Follow-up or whenever the last interview with R took place;

3. calling contact numbers, such as parent numbers, to obtain a number for respondents when the first number called was incorrect and Directory Assistance did not provide a new number;

4. performing locating steps to find a respondent, and

5. performing refusal conversion.

TNMS was programmed to route cases to different locations within itself. As shown above, the TNMS locations were organized according to data collection task: contacting, interviewing, locating, and refusal conversion. Automated TNMS procedures were established to route cases from location to location depending on outcome selected. These procedures were being used extensively for the first time on HS&B and were successful. The change to hard copy sample management was initiated when application of the extensive TNMS rules to few cases slowed down case delivery. This change allowed interviewing to continue in an efficient manner regardless of routing procedure problems.

5.5.4 On-Line Coding

For the fourth follow-up, interviewers performed on-line coding tasks. Interviewers were trained to code respondents' postsecondary institutions, occupations, and industries in which they worked. In training, interviewers were required to successfully complete exercises in each type of coding, and were allowed to practice coding prior to beginning data collection. Interviewers were also trained to record respondents' verbatim descriptions of industry and occupation for researchers who wish to code at a more detailed level.

Industry Coding. The coding scheme for industry used on the main survey was a simplified version of the U.S. Department of Commerce, Bureau of the Census Classified Index of Industries and Occupations 1970, and the U.S. Department of Commerce, Bureau of the Census Alphabetical Index of Industries and Occupations 1970, which had been used on previous rounds of HS&B. To simplify the coding task, the major headings were used as the coding categories with one exception: Manufacturing was split into Durable and Nondurable Goods. For researchers who wish to code at a more detailed level, the entire verbatim response is also reported in the data file.

In the main survey, interviewers reported little difficulty using the coding program for industry coding. AutoQuest was programmed to allow interviewers to search for an industry by entering a search string, usually a key word in the respondent's description. The interviewer was shown a list of all codes that contained the search string. The interviewer queried the respondent about the possible choices and coded based on the respondent's input. This technique allowed for respondent input during the coding process and improved the coding accuracy rate. Occupation Coding. Interviewers used the same process in AutoQuest to code occupation as described above in industry coding. The coding scheme for occupation coding used on the main survey was adapted from the HS&B Third Follow-Up Questionnaire, which asked the respondent to write the name of the job or occupation that he or she expected to have at 30 years of age. If the respondent wasn't sure, he or she was instructed to write in one best guess at what the expected job or occupation might be, and was then asked to select from the 18 categories listed the job that came closest.

5.5.5 Postsecondary Institution (FICE) Coding

During the main survey, interviewers coded respondents' postsecondary institutions on-line, using the Federal Interagency Committee on Education (FICE) list of institutions that was augmented during previous rounds. The look-up tables enabled the user to complete a search by entering parameters such as name, city or state of the institution attended by the respondent. NCES provided an updated version of the FICE table offering somewhat more consistency in the manner in which postsecondary institutions were listed, and making the task of finding the institution on the table easier for interviewers. Also, the FICE table data were printed and distributed to interviewers during additional FICE training to help interviewers understand the nature of the entries in the table. During training, interviewers were expected to successfully complete an exercise in FICE coding and were allowed to practice FICE coding before beginning data collection.

If during the interview an interviewer was unable to ascertain the FICE code, he/she collected the institution name, city and state. A coding specialist experienced with FICE coding reviewed the text in an effort to code the institution. Institutions not on the augmented list were assigned dummy FICE codes and added to the list.

5.5.6 Monitoring

Telephone Center operations were monitored by CATI supervisors to ensure consistent high-quality data throughout the field period. Both the voice and computer screen portions of all interviewer and locator activities were monitored on-line. Interviewers could also be monitored remotely.

There were two systems developed for monitoring. The first was used to draw a statistically valid sample of all shop activities prior to the start of each day's work. The program randomly selected a sample from among the stations used by both locators and interviewers and assigned random start times to the selected stations. Monitors were given this schedule and instructed to monitor whatever activities took place in the 15 minutes following the start time.

The second program was designed to capture monitoring information, which was collected on paper forms as the monitoring session progressed. The program collected session start and stop times, monitor ID, the ID of the interviewer being monitored, and the status of the station.

The next screen captured the activity currently being monitored: interviewing, locating, or gaining access and cooperation. A final program screen allowed entry of the item identifier and error code for each item on which an error or deviation occurred. The monitor could then append a note indicating the type of error that was observed.

Statistical control charts were employed to monitor whether or not the telephone center error rate was statistically in control. Only on one occasion were activities not in control. An investigation determined that one supervisor on that day used different criteria than other supervisors in judging deviations and errors; the supervisor was subsequently retrained.

5.6 Data Processing

Data processing activities span the entire length of each of the HS&B surveys, beginning with pretest activities, continuing with maintenance of the respondent locator database, and concluding with machine editing and the preparation of public use data tapes. Data processing activities in the base year and in the first through fourth follow-ups are discussed together in this section.

5.6.1 Maintenance of Longitudinal Locator Databases

The locator database maintains the most up-to-date name and address information available for each sample member as well as information from previous waves. During each wave, respondents have completed a locator page that requests their names and addresses, their spouses' names, their parents' names and address(es), and the names, addresses, and relationships of two other people who are likely to stay informed of the respondents' whereabouts. The locator page also requests information regarding respondent birth date, sex, and social security number. For the fourth follow-up the "locator page" was included in the CATI instrument. To ensure confidentiality, all locating information is stored in secure files that are separate from the questionnaire data.

Since five surveys have been completed and since birth date and sex are also provided elsewhere in each questionnaire, several independent sources of locating and identifying information are generally available. This information is necessary for locating hard-to-find respondents, verifying that a given ID number refers to the same individual across waves, and constructing corroborated birthdate and sex composites (BIRTHMO, BIRTHDAY, BIRTHYR, and SEXCOMP).

5.6.2 Receipt Control Procedures

For the first three waves (base year through second follow-up), the NORC Automated Survey System (NASS) was used to track survey activities. This system houses a data file for each school in the base year and first follow-up surveys and for each cohort in all waves; the respondent ID number; disposition codes; and other information. During the base year, the school NASS file was used to generate weekly summary reports that tracked refusal rates and patterns, completed survey days or delays, and the receipt of school-level documents (i.e., school questionnaires). NASS also generated customized calendars of scheduled school survey days for each NORC Survey Representative.

For the third follow-up, the Survey Management System (SMS) was used, which is functionally equivalent to NASS but has some additional capabilities. Because it interfaces with CADE, it could update internal dispositions automatically and generate reports on the progress of the documents as they were processed.

During the base year and each follow-up, weekly summary reports on the receipt of sophomore and senior questionnaires were produced. Data control disposition codes were added to the NASS/SMS files, making it possible to track the internal movement of instruments through mail receipt, editing, data retrieval, validation, CADE, and shipment for optical scanning. The respondent-level NASS/SMS files were linked with the longitudinal locator database to produce interviewer assignment logs, to trace nonrespondents as of a certain date, and to produce reminder postcards. The NASS/SMS also generated the transmittal materials for shipping the prepared instruments to the optical scanning subcontractor.

At the end of each data collection period for the first and second follow-ups, a reconciliation between the files provided an accurate count of the number of survey participants and documents received. The reconciliation used three types of checks: check digits derived from a fixed mathematical formula that easily identified misread or miscopied student ID numbers; a comparison of the respondent's birth date, sex, and other identifying information against base year and first follow-up data; and a comparison of field transmittal forms against what the NASS records indicated had been returned from the field. All discrepancies were reported for review and resolution.

Reconciliation for the third follow-up was somewhat different due to the fact that data were converted by both CADE and optical scanning. In order to reconcile third follow-up data with prior waves, every ID was checked against a master list before data were entered in CADE. Once CADE and scanning operations were complete, NORC matched the Questar tape with the CADE data file and reported any discrepancies. Each case was examined individually to determine whether an ID had been miskeyed. Although all questionnaires had been preprinted with the ID for optical scanning IDs were entered by hand for questionnaires that had been remailed and questionnaires that had been administered by field interviewers. Consequently, errors in IDs were possible, and all discrepancies were reported and resolved.

Instrument control for the fourth follow-up was managed through the TNMS (5.5.3) for telephone center cases and through the Field Management System (FMS) for cases sent to field interviewers. See Section 6.6 for receipt control procedures for fourth follow-up transcripts.

5.6.3 Optical Scanning

Prior to the fourth follow-up, the student questionnaires were optically scanned using equipment that read darkened ovals or marks on the page. For each survey, the scanning subcontractor conducted extensive tests and checks of the machine's ability to correctly read the darkened ovals. Adjustments were made to the mark-sense threshold as required. Finally, questionnaires were marked up and scanned. The results were then compared with hard copy to verify that satisfactory data conversion was being achieved.

In the base year, student instruments were limited to two versions (one per cohort) and the instruments contained only one logical branch or skip sequence for respondents to follow. Because of this simplicity, it was efficient for the optical scanning contractor to perform the critical item edit and convert blank fields to missing value codes at the time of completing the data conversion. The conversion of blanks to missing values was done according to instructions from NORC.

The optical scanning contractor for the first three waves was National Computer Systems (NCS). (In the base year the company was called Westinghouse Learning Corporation, and during the first follow-up, its name was changed to Westinghouse Information Services.) For the third follow-up, the scanning contractor was Questar Data Systems Inc.

For the first three surveys, NCS created separate data files for the two cohorts. To check the accuracy of data conversion, NORC conducted an audit of a sample of cases, comparing the scanned and machine-edited data files with the hard-copy questionnaires.

In the third follow-up, there was a single instrument for both cohorts. As discussed earlier, a portion of the instrument was designed for CADE, while the rest was prepared for optical scanning. All major skip items and all critical items were entered in CADE. Missing values were converted to blanks. During machine editing at NORC, blanks were changed to missing value codes. Because there was only one instrument in the third follow-up, only one data file was prepared for the two cohorts. To check the accuracy of data conversion, NORC audited a sample of 100 cases, and final data were compared item by item to hard-copy questionnaires and procedures were modified until accuracy was attained.

The fourth follow-up did not use optical scanning to capture data; the CATI system captured the data at the time of the interview. A CADE program was designed to enter and code transcript data.

5.6.4 Machine Editing

In the base year, machine editing was limited to examining each data field for out-of-range values. Very few stray codes were discovered; appropriate missing value codes were assigned to these fields. As noted in the section on optical scanning, base year questionnaires were designed so that only one explicit skip instruction appeared in the senior questionnaire (seniors not going on to college did not complete the last section on college education). There were no skip instructions in the sophomore questionnaire. Where two or more questions were related, the items following an implicit screening or filter question contained response options for those who were screened out by the filter question. No inter-item consistency checks were carried out on base year data files between the implicit filter questions and the related (dependent) items.

In the first and second follow-ups, several sections in the questionnaire required respondents to follow skip instructions. A case by-case inspection of logical inconsistencies and stray codes was impractical due to the sheer number of cases and the fact that the pages of the questionnaires had been cut apart in preparation for data entry by optical scanning. Consequently, programs were written to automatically perform the inter-item machine-edit checks. The tasks performed included: resolving inconsistencies between filter and dependent questions, supplying the appropriate missing data codes for questions left blank, detecting illegal codes and converting them to missing data codes, and generating a report on the quality of the data as measured by the incidence of correctly and incorrectly answered fields and correctly or incorrectly skipped fields.

Inconsistencies between filter and dependent questions were resolved in consultation with NCES staff. In most instances, dependent questions that conflicted with the skip instructions of a filter question contained data that, although possibly valid, were superfluous. For instance, respondents sometimes indicate 'no' to the filter item and then continue to answer "no" to subsequent dependent questions. Data retrieval verified that filter questions were generally answered correctly, and dependent questions that should have been skipped were often inadvertently answered because they seemed to apply. During the machine-editing process, inappropriate responses were expunged by turning them into blanks.

After improperly answered questions were converted to blanks, the student data were passed to a program that supplied the appropriate missing-data codes for blank questions. The program converted questions left blank according to several criteria. If a previous question had been answered in a way that required that the current question be skipped, a "legitimate skip" code was supplied. If not, a "missing data" code was supplied, except in the case of critical questions, which were flagged during data preparation and attempts were made to obtain the information by telephone. If the respondent specifically refused to answer a question during the call-back, a special scannable oval was marked. Critical questions marked in this way were assigned a special missing data code of "refused." Otherwise, critical questions were treated in the same manner as others. Finally, additional missing value codes for multiple-coded questions were supplied by the scanner.

Detection of out-of-range codes was completed during scanning for all questions except those permitting an open-ended response. For the hand-coded, open-ended questions (such as the three-digit occupation and industry codes and the six-digit college and field-of-study codes), the data were matched by computer against lists of valid codes, and invalid codes were converted to missing values. The numbers of invalid codes detected were negligible.

For measuring data quality, the machine-edit programs produced bar graphs that displayed the frequencies for the different situations recognized by the programs: questions properly answered, questions properly skipped (the "legitimate skip" code), questions skipped in error ("missing data" code), and questions answered in error.

The treatment of inappropriately answered items (i.e., those a respondent was instructed to skip) relied on the results of the critical item editing procedure. With only one or two exceptions, screening or filter questions were designated as critical items. When respondents were inconsistent in answering these items, either by responding to items they were instructed to skip or by failing to answer the dependent questions related to a filter item, the case was classified as an edit failure. As discussed in section 5.3.3, telephone calls were used to obtain responses to items skipped in error. The results of these calls demonstrated unambiguously that inappropriate answers to filter-dependent items were universally caused by respondents' failure to comply with the routing instructions of the filter questions. Rather than skipping to the designated target question to resume their responses, these individuals attempted to answer each filter-dependent question that appeared to offer a reasonably suitable response category. On the strength of these findings, all filter-dependent responses entered in error were converted to the proper missing data values (i.e., the "legitimate skip" code).

During the third follow-up, CADE carried out many of the steps that normally occur during machine editing. The system enforced skip patterns, range checking, and appropriate use of reserved codes, which allowed operators to deal with problems or inconsistencies when they still had the document in hand and consequently had the most information available (see Section 5.4.2).

For the items that were scanned, the same machine editing steps as those used in prior follow-ups were implemented. Since most of the filter questions were CADE designated items, there were few filter-dependent inconsistencies to be handled in machine editing.

For the fourth follow-up, machine editing was replaced by the interactive edit capabilities of the CATI system. During the interview, interviewers were warned of out-of-range responses and resolved these types of problems with the respondent. (See section 5.5.1.)

5.6.5 Data File Preparation

In the base year, data for the two cohorts were combined into a single data set. To facilitate this, NORC reformatted the tape so that questions identical in the two versions of the questionnaire occupied the same tape positions in each record. In general, the data for both cohorts followed the order of the senior questionnaire. Items unique to the sophomore instrument were interspersed among the senior items so that sophomore data appeared in about the same order as in the questionnaire. Also, whenever necessary, the sophomore response category values were recoded to match those for the senior cohort.

Data for the first follow-up were merged with base year data, and a merged data set was created for each cohort and placed on its own tape. After the second follow-up was completed, these data were merged with the base year and first follow-up files. Similarly, third follow-up data were merged with base year and first and second follow-ups.

For the fourth follow-up, three data files were created: a student file containing student-level data collected from all five rounds, a transcript-level file and a course-level file.

6. SOPHOMORE COHORT POSTSECONDARY EDUCATION TRANSCRIPT STUDY

6.1 Scope of the Postsecondary Education Transcript Studies

Although the HS&B follow-up surveys have collected longitudinal data on postsecondary educational activities of sample members, the kinds and quantity of information collected on course-taking patterns and on grades, credits, and credentials earned has necessarily been limited by the survey methodology, and by respondents' ability to recall and accurately report the details of their educational experiences.

To overcome these weaknesses and to provide a rich resource for the future analysis of occupational and career outcomes, transcript information was abstracted and coded. Thus, they can be merged with questionnaire data and other records data (e.g., information from students financial aid records) to support powerful quantitative analyses of the impacts of postsecondary education.

Data files created for the transcript study include detailed information about program enrollments, periods of study, fields of study pursued, specific courses taken, and credentials earned. In addition to providing a data resource for the analysis of educational activities and their impacts, the transcript data may be used as an objective standard against which students' self- reports may be compared and evaluated, thus guiding the design of future studies.

Transcript requests for the Sophomore Cohort Postsecondary Transcript Study were made for the subset of the sophomore cohort who reported in the follow-up survey that they had attended a postsecondary institution (see Sample Design and Implementation below).

6.2 Transcript Data Collection

To be included in the study, an institution had to be on the then current IPEDS list. Using this criterion, 872 institutions reported by the sophomore cohort sample members were included from the transcript study.

Preparations for collecting and processing all other transcripts included three major steps:

1. Extracting information concerning each unique instance of postsecondary institution attendance by sophomore cohort members from HS&B follow-up survey data files and sorting this information by institution name and identification number. These data were used to generate the printed lists of students sent to registrars and other institution administrators to request transcripts.

2. Materials production - Constructing up-to-date address files for all postsecondary institutions reported by sample members, and developing letters, forms and other materials to be sent to institution administrators explaining the purposes of the study, the legal authority under which the study was being conducted, and procedures for protecting the confidentiality of research subjects.

3. Obtaining the endorsement and support of a broad spectrum of professional organizations engaged in research about and representing the interests of postsecondary institutions.

6.3 Transcript Data Collection Objectives

The principal objective of the study was to obtain all transcripts for sample members who reported attending postsecondary institutions. In addition, course catalogs and other related publications were requested from these institutions to facilitate the accurate and consistent coding of information about programs or fields of study, course titles, earned credits, grades, degrees or other credentials, and academic terms or other measures of enrollment duration.

A secondary objective of the transcript study was to validate self- reporting by sample members of postsecondary institutional enrollment. Thus, transcripts were requested from each institution reported in follow-up questionnaires, even if there was evidence that the respondent might not have completed the term of study or the requirements for credit. As indicated by the results described below, in a small percentage of cases the institutions reported that the respondent either never actually attended classes at the named institution, or else dropped out of classes before completing enough work to justify creation of a formal record.

6.4 Mailout of Transcript Request to Institutions

During the week of February 22, 1993, packets of transcript survey materials were mailed to the postsecondary institutions. The mailing was timed to arrive at registrars' or other administrative offices at a time of low level of activity for the administrative staff.

Each transcript request package contained the following, of which examples are provided in Appendix A:

1. a list of postsecondary organizations endorsing the transcript study

2. a letter to the Registrar from the NORC High School and Beyond Project Director

3. an endorsement from the American Association of Collegiate Registrars and Admissions Officers (AACRAO)

4. a letter from the Commissioner of the National Center for Education Statistics authorizing NORC to conduct the study on behalf of the Secretary of Education

5. an excerpt from the Family Educational Rights and Privacy Act (FERPA) indicating the legal authorization under which the request for records was made (copy not in appendix)

6. a brief description of NCES's National Education Longitudinal Studies program

7. general instructions for participation in the study

8. a computer-generated list of students for whom transcripts were being requested (copy not in appendix)

9. a label to affix to each transcript to link the correct transcripts to HS&B files (copy not included in the appendix)

10. an invoice form for transcript reimbursement (copy not included in the appendix)

11. a prepaid address label for transcript shipment (copy not included in the appendix)

Telephone follow-up of non-responding institutions began in early April when transcripts had been received from about 47 percent of the institutions.

6.5 Data Collection Results

To a great degree the success of the transcript study hinged on the cooperation of registrars and other administrators to whom transcript requests were sent. Despite the fact that study materials fully explained the legal basis for the requests for the information, institution officials had the right to decline to cooperate. Most officials supported the objectives of the study, however, and were complete in their responses. Even so, other logistical obstacles had to be overcome. A number of institutions, all in the vocational and proprietary sector, had either permanently closed, or indicated only kept records for a limited amount of time (usually five years). Other institutions relocated, changed their names, or merged with other institutions necessitating extensive tracing efforts in order to deliver requests to appropriate offices, and complicating the task of locating specific student records. In other sections we describe the response rates at three levels - the institution, the individual transcript (instance of attendance), and the student (for whom more than one transcript may have been requested).

6.5.1 The Institution-Level Response Rate

Transcript requests for HS&B students were sent to a great variety of postsecondary institution types, including small and large private vocational and proprietary institutions as well as traditional degree-granting institutions of higher education such as 2- and 4-year colleges and universities with the full range of graduate and professional programs. Identical materials and procedures were used in the collection of transcripts for all types of institutions. However, as shown in Table 6.1, more non- vocational institutions (e.g., colleges and universities) participated in the study more frequently than did their vocational counterparts (e.g. trade and technical institutions). The participation rates shown in the table are the simple percentages of institutions in each sector that returned at least one transcript. No attempt was made in this table to adjust either for the number of transcripts requested or for the possibility that all transcripts were requested for students who did not actually attend the institution. (Transcripts were classified as "out-of-scope" as a result of information returned by institution personnel indicating that the individuals for whom transcripts were requested never attended their institutions or did not complete enough work to generate a formal record).

Only 50.4% of the private, for profit institutions returned any transcripts. This institution type tended to be less cooperative (see Exhibit 6.3) than the other institution types. Almost as important is the higher incidence of not being able to find or supply records for students who attended the institutions. This may be attributed to the tendency not to keep student records beyond 5 years. The sector, however, constituted only 22.3% of the eligible institutions and roughly 6.4% of all transcript requests.

Table 6.1--Response rates to the HS&B postsecondary education transcript study by institution type

	Response rate (Percent)	Number of institutions in sector
Private, for-profit	50.4	752
Private, not-for-profit	75.5	151
Public, less-than-2-year	75.3	271
Public 2-year	93.5	800
Private, not-for-profit 4-year	92.6	809
Public 4-year	95.1	555
Unknown	6.5	32
Total	80.8	3370

Table 6.2--Transcript dispositions: out-of-scope and in scope by institution type by percentage and raw numbers

	Out of scope	In scope	Total
Private, for-profit	21.7% (269)	78.3% (969)	100.0% (1238)
Private, not-for-profit less-than 4-year	10.1% (29)	89.9% (259)	100.0% (288)
Public, less than 2-year	15.9% (119)	84.1% (631)	100.0% (750)
Public 2-year	7.4% (402)	92.6% (5007)	100.0% (5409)
Private, not-for-profit 4-year	8.3% (254)	91.7% (2825)	100.0% (3079)
Public 4-year	7.3% (423)	92.7% (5380)	100.0% (5803)
Unknown	2.9% (1)	97.1% (33)	100.0% (34)
Total	9.0% (1497)	91.0% (15104)	100.0% (16601)

Table 6.3--In scope transcript dispositions: by institution type

	Received	School refused/ non-response	Lost or destroyed	School closed	Total
Private, for-profit	59.8%	35.3%	1.3%	3.6%	100%
	(580)	(342)	(12)	(35)	(969)
Private, not-for-profit less-than-4-year	83.0%	15.8%	0.4%	0.8%	100%
	(215)	(41)	(1)	(2)	(259)
Public, less than 2-year	80.8%	18.1%	0.8%	0.3%	100%
	(510)	(114)	(5)	(2)	(631)
Public 2-year	91.3%	8.5%	0.2%	0.0%	100%
	(4570)	(425)	(12)	(0)	(5007)
Private, not-for-profit 4-year	93.5%	6.2%	0.1%	0.1%	100%
	(2642)	(176)	(4)	(3)	(2825)
Public 4-year	94.5%	5.3%	0.2%	0.0%	100%
	(5083)	(287)	(9)	(1)	(5380)
Unknown	9.1%	90.9%	0.0%	0.0%	100%
	(3)	(30)	(0)	(0)	(33)
Total	90.1%	9.4%	0.3%	0.3%	100%
	(13603)	(1415)	(43)	(43)	(15104)

6.5.2 Transcript-Level Response Rate

Requested transcripts are defined as:

17,619	reported by students
- 1,018	transcripts from out-of-scope institutions (see 6.5.1)

16,601	transcripts requested

Transcript response rates are calculated as ratios of the number of transcripts received to the transcripts requested. Transcripts were classified as "out-of-scope" as a result of information returned by institution personnel indicating that the individuals for whom transcripts were requested never attended their institutions (or did not complete enough work to generate a formal record). These transcripts have been treated as outside the population of events being studied rather than as "missing observations." Given this response rate definition, 90.1% of eligible transcripts were processed (see table 6.3).

Table 6.2 shows the magnitude of cases classified as out-of-scope (9% overall). The percentage out-of-scope is lowest (7.3 to 8.3 percent) among public and private 4-year institutions and public, 2-year institutions. The percentage increases to 10.1 percent for private, less than 4 year institutions and to 15.9 percent for public, less than 2 year institutions. It reaches its highest level (21.7 percent) for private, for-profit institutions.

Since the initial list of instances of institution attendance was created using survey responses to the HS&B third and fourth follow-up surveys, these results create inconsistencies between the questionnaire data files and the postsecondary transcript study data file. The discrepancy between student- reported postsecondary attendance and the evidence in institution records is substantial, and the decision to consider these instances as out-of-scope was not taken lightly. It is important to note that this status code was only assigned to cases when institution officials confirmed in writing their conclusion that the named student did not attend their institution. Administrators had considerable information about each student named on a transcript request form, including full names, alternative names such as maiden names, social security numbers, dates of birth, and approximate dates of enrollment. In addition, there was considerable evidence in the materials returned and telephone calls to NORC that institution personnel had conducted thorough searches for records, and often had cross-checked their results with admissions offices and financial aid offices. We therefore believe that there is little or no classification error in this status code.

One interpretation of this outcome is that HS&B respondents over reported instances of postsecondary attendance. If so, researchers analyzing postsecondary schooling using only the survey data would overestimate the extent of this activity. Furthermore, the true discrepancy may be even greater than that estimated by these results. For a portion of the cases in the "School Refused/No Response" category of Table 6.3, neither transcripts nor any other information about the students' status was returned. In the absence of specific information to the contrary, these cases have been treated as missing instances of attendance, and therefore within the scope of the population of interest. It is reasonable to expect that if information had been obtained for these cases, some portion would have been declared as errors in reported attendance.

The fact that the rate of "Never Attended" classifications is higher among proprietary and public, less than 2-year institutions is consistent with descriptions of the incidence of last-minute withdrawals and dropout rates at these institutions. However, the evidence is strong enough to rule out alternative interpretations. One reasonable alternate possibility is that some of these instances of reported attendance result from errors in the coding of institutions. For the first time the FICE coding task was handled on-line by interviewers during the CATI interview. Coding of institutions was previously a task handled by coding specialists after the interview.

On the one hand, "post-coding" does not allow for probing to clarify information about the institution. On the other hand, on-line coding has its own deficiencies. For example, some institutions had more than one FICE code and rules/guidelines for choosing codes evolved as the data collection period progressed.

Conceivably, respondents may have in fact attended a postsecondary institution but the name and FICE reported is incorrect. After these out-of- scope transcripts are excluded, Table 6.3 shows data collection results at the level of the individual transcript for the total sample, and separately for each of the six types of postsecondary institution.

As can be seen in Table 6.3, reasons for non-return of transcripts varied among institution types. Institution refusal and non-response accounted for 9.4 percent of missing transcripts. Confirmed institutions closing affected only 0.3 percent of transcripts. Overall, 0.3 percent of transcripts were not available because records had been lost or destroyed, or transcript records were only available for the most recent years.

6.5.3 Student-Level Data Collection Results

Transcripts were sought for 9,064 HS&B 1980 Sophomore members who reported attending postsecondary institutions. Reports of postsecondary attendance were obtained from HS&B third and fourth follow-up survey questionnaire responses. Table 6.4 presents distributions of the number of transcripts received for each student. Excluding the out-of-scope cases, one or more transcripts were obtained for 93.2 percent. A single transcript was received for 52.0 percent. Two transcripts were processed for 28.5 percent and three or more transcripts were obtained for 12.7 percent.

Table 6.4--Number of transcripts received: HS&B postsecondary education transcript study

	Number of respondents	Percent
None	617	6.8
One	4,714	52.0
Two	2,587	28.5
Three	916	10.1
Four	192	2.1
Five or more	38	0.4
Total	9,064	100.0

In addition to collecting multiple transcripts per student, many transcripts contained information about credits transferred from other institutions. Transfer credits were specially flagged in the data files to assist researchers in avoiding double-counting of earned academic credits by those who attended more than one institution. Transfer credits for these individuals have been documented in their transcript records. The variables TRNSFERS on the student-level record and TRANFERT on the transcript-level record in the data files identify individuals and transcripts containing transfer credits.

6.6 Data Preparation

6.6.1 Data Preparation Objectives

The diversity in structure and contents that exists among the transcript records reflects the great variability among the institutions from which they were obtained. Although transcripts from public and private 2-year and 4-year colleges were generally similar with respect to the data they contained, they nonetheless differed in their physical layout and in the terminology used. The apparent similarities in many transcripts give way to countless differences in the ways in which academic progress is measured and recorded. This is especially true of course grades and credits.

The variability across institutions in the details of transcript information defies any simple aggregation or homogenization. Virtually any element in an academic transcript, including such seemingly straightforward items as course titles, may be subject to highly particularized local conventions whose logic may be independent of or even contravene, common practices. For example, it is not uncommon to find courses in English composition merged with other content and carrying formal names suggesting that they belong in the social science curriculum. Such instances, by no means rare, were resolved by Computer Assisted Data Entry (CADE) staff, who consulted program-of-study catalogs and descriptions of courses obtained from the postsecondary institutions.

In preparing the data for conversion to standardized, machine-readable data files, NORC's approach was twofold. The first step was to impose a common structure and organization on the transcript information to enable us to preserve the actual information contained in the original documents. The second step was to assign numeric codes to certain elements such as degrees and credentials earned, major and minor fields of study, and titles of courses taken using a common coding frame. Either the original data or the coded values can be accessed by researchers and used as they see fit. The coded values were also utilized to create variables that shared a common metric. This was done to ease comparisons of data collected from different institutions. More discussion of this issue can be found in section 6.9.

6.6.2 Data Organization

Transcript data were organized into the three-level hierarchy consisting of data at the student, transcript, and course levels. At least one student- level and one transcript-level record is provided for each sample member who reported postsecondary attendance. Therefore, there are student transcript records even if the institution reported that the individual had never attended, or had withdrawn before establishing a formal record. Records in this category are flagged with a special disposition code.

Student-level data refer to general information about the respondent's educational career such as institutions attended, degrees attempted and attained, highest degree attained, and dates of attainment. All records are assigned case ID codes, allowing merger with other files (transcript and course), questionnaire data from the HS&B base year and follow-up surveys.

Transcript-level records contain data pertaining to a student's academic record at a single institution, including the institutional ID code (FICE code), degree(s) or other credentials conferred with accompanying dates, major and minor field(s) of study, and the student's cumulative grade point average (GPA).

Course-level records store the data for each course taken by a student. The formal title of the course was entered verbatim from the transcript, then assigned one of the codes contained in the publication: A College Course Map Taxonomy and Transcript Data,<1> or CCM. An additional code was reserved to indicate lump-sum transfer course credit. Also entered were credits attempted and the grade received by the student for each course, term type (e.g., semester, quarter) and term dates.

6.7 Computer Assisted Data Entry (CADE) and Coding

The 1993 HS&B Postsecondary Transcripts Study had two phases of data processing: recoding 1987 Transcript Study data and the data entry/coding of newly acquired transcripts. Data from both phases were processed using a CADE program and were integrated prior to final delivery of the data.

During Phase 1, a team of 5 coders with college credentials were hired to recode the 1987 transcript course data (which was originally coded in a different format from the CCM). Data were loaded into a sequential query language (SQL)-based coding program, and coders used a menu-driven coding engine (prepared by NCES) to display possible course codes and descriptive summaries of the CCM codes. Through a series of commands, coders could either choose from a list of possible codes or input the CCM code directly. The CADE format enforced the predetermined set of CCM codes and field of study codes. Other value limitations made it impossible for CADE operators to enter an illegitimate transcript ID.

Through recoding, six-digit CCM codes replaced the 2-digit codes applied in 1987 from A Classification of Instructional Programs.<2> Staff recoded major and minor fields of study as well.

The Phase 2 portion of the study required the abstracting, data capturing, and coding of data from thousands of newly acquired transcripts that varied greatly in appearance and content. As transcripts were received, data entry clerks, selected from the existing staff of CADE operators, were trained to abstract and key the data into CADE screens. In addition to capturing the data, data entry clerks determined the institutions' grading scales and term types and identified transfer courses.

The captured data were then loaded into the coding program which displayed the structured transcript information online. The coding clerks assessed the data and applied codes. Clerks could refer to the hard copy transcripts and course catalogs, as necessary, but for the most part, they worked from CRT screens as they entered the codes.

During training and production, emphasis was placed on "coding in context," which meant applying codes based not only on the course name and the department offering the course, but also on 1) the point at which the student took the course as he/she progressed through the curriculum, 2) related coursework taken, and 3) the number of credits earned. Based on these factors, a coder might apply a code for a higher level course even though a simple reading of the course title suggested an entry-level course code or vice versa.

6.8 Data Quality Management

Quality control of transcript coding was introduced and maintained through a combination of procedures: error prevention features within the CADE program and double entry of some of the transcripts followed by review of any discrepancies between the first and second coder. This verification procedure enabled management to better assess the degree of agreement among coders. Verifier re-entry of transcripts involved 1,165 transcripts, or 8.6 percent of the transcripts processed. In addition, the discrepancies were discussed among the coding staff and, if necessary any ambiguities were brought the attention of NCES in a regularly scheduled biweekly meeting. These phone meetings were attended by the entire coding staff who had an opportunity to discuss courses for which they were unsure of the appropriate code.

All uncodable course were also sent to NCES for resolution by the author of the CCM. In order to code in context, NCES received all coursework and field of study information from the transcript in an electronic file. Once NCES resolved the issues, the file was returned and the new codes were added to the existing data.

The CADE program itself screened for error in three ways. Through the use of preloaded data, the program prevented entry of incorrect identification data (i.e., institution FICE codes, student ID numbers, and combinations of institutions and students). Furthermore, each data field was programmed to disallow entry of illogical or otherwise incorrect data. For example a data entry clerk was automatically prevented from entering a letter grade for a course if numerical grading system had been specified. Further, it was not possible to enter a non-existent code.

As unanticipated problems arose during the CADE period, a policy decisions protocol was followed. All questions and other issues were directed to project management and NCES staff for assessment and final coding decisions. The resulting decisions were routinely distributed to the CADE operators to be added to their coding manuals.

6.9. Data Processing

Data Processing activities began with the development of a document control system that could monitor Phase 1 and Phase 2 activities. Development of the CADE coding system followed. While staff recoded the 1987 transcript data, the High School and Beyond fourth follow-up CATI data were analyzed to determine which transcripts were to be requested in 1993. Once identified, customized transcript request packets were prepared with the aid of programmers.

After all transcript data were converted to machine-readable form, data were uploaded from the local area network (LAN) to mainframe facilities to expedite the processing. CATI-transcript record linkages were created by reconciling the transcript records with the fourth follow-up CATI data. At this point new variables were created to help analysts compare data collected from different institutions. As noted previously, institutions use a wide range of formats and scales when reporting such items as course credits and grades.

Variables were created to standardize grades and grade point averages, credit hours, course types and major and minor fields of study. Further information about these items can be found in the codebook that is included with the data files. Analysts are advised to thoroughly review these items to determine if they meet their analytical needs.

Transcript weights were developed, and all transcript related data were then restructured into two main transcript files containing transcript- and course-level data. Other transcript variables were appended to the student-level data.

Finally, program control cards were generated to permit the construction of analysis files using either SPSS or SAS.

END NOTES

<1> A College Course Map Taxonomy and Transcript Data (Adelman, Clifford; Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education 1990)

<2> A Classification of Instructional Programs (Maliz, G.S., et al.; Washington, D.C.: National Center for Education Statistics, U.S. Department of Education 1981)

7.0 DATA QUALITY

Several sources are available to analyze the quality of the HS&B fourth follow-up CATI and transcript data. First, we will evaluate the CATI data by examining data collected through the monitoring of interviews. Second, we will evaluate missing response rates and patterns by looking at both third and fourth follow-up survey data. Third, we will evaluate the consistency of responses between the third and fourth follow-up data, specifically examining marital status and race. Finally, we will examine some proprietary institution non-response issues and possible bias introduced into the transcript data.

7.1 Monitoring

During the HS&B fourth follow-up, NORC used a monitoring system designed to obtain a statistical sample of interviewer activity. A supervisor was given a schedule each day of randomly selected times and interviewing stations. At the appointed time, supervisors monitored all activity occurring at the station between designated start and stop times.

Overall, approximately 1% of interviewing (including locating) was monitored. Most of the monitoring was done between March and June and was roughly proportional to the level of activity in the telephone center. By month, the total minutes of monitoring were:

February	514
March	2,624
April	7,220
May	3,105
June	2,638
July	976
August	956
September	512

This monitoring had two purposes. First, the monitoring data was used to determine the overall quality of the data collected by the interviewers. Second, the monitoring data was used to improve the interviewing by eliminating preventable errors. Thus, the interviewers could receive feedback on their interviewing skills as the study continued.

Mistakes were defined as any significant departure from the script, and were divided into two categories: deviations and errors. Errors were defined as departures that could adversely affect the quality of the data, such as asking of a question in a biased way. Deviations, on the other hand, were defined as less harmful departures, such as substitutions of words that might be better understood by the respondents. In assessing data quality, we look only at errors below.

The activities monitored were divided into three types: gaining cooperation, the introduction questionnaire, and the main questionnaire. Since gaining cooperation was often intermixed with the introduction, the distinction between these categories is not perfect.

The overall error rate was 0.025 errors/minute of monitoring (465 errors in 18545 minutes). This is about 1 error every 40 minutes. No errors were detected in 6019 minutes classified as the Introduction. The error rate for Gaining Cooperation was a very small 0.003 errors per minute (20 errors in 6218 minutes), which is about 1 error every 5.5 hours. Also, there were no errors detected after April for gaining cooperation. The error rate for the main questionnaire was 0.072, however (445 errors in 6308 minutes), which is about 1 error every 14 minutes, or about 2 per completed interview.

Table 7.1 shows monthly error rates for each of these four components.

Table 7.1--Monthly error rates for each monitoring component, in errors per minute

```
0.10 |
     |
0.09 |          Q
     |  Q
0.08 |
     |                                          Q
0.07 |                          Q
     |                  Q               Q
0.06 |
     |
0.05 |
     |
0.04 |                                                          Q
     |  T       T       T
0.03 |
     |
0.02 |
     |                          T       T
0.01 |                  G                       T               T
     |          G
0.00 |  IG      I       I       IG      IG      IG      QTIG    IG
      ---------------------------------------------------------------
      Feb     Mar      Apr     May      Jun     Jul      Aug     Sep
```

(Key: Q = Main Questionnaire, T = Overall, G = Gaining Cooperation, and I=Introduction):

There seems to be a general decline in the overall error rate. This seems to be due to the decline in the main questionnaire error rate. However, there is a confounding factor. As the study continued, the percentage of interviewing monitored that consisted of the main questionnaire decreases. To understand the overall error rate, we really only need to look at the main questionnaire monitoring data because there are very few errors in the other two categories.

Besides examining the monthly error (shown above), the data were looked at 3 different ways: weekly, daily (after smoothing by adding previous and subsequent days to each day), and using time periods based on minutes monitored (i.e., the 1st 50 minutes monitored, the

2nd 50 minutes monitored, etc.). All four analyses showed that there was a decline at the very end of the interviewing, during August and September.

However, it is unclear whether there is any decline before these months. The daily and monthly analyses show some evidence that the error rate declined when the bulk of the interviewing began (April). However, this decline is small compared to the decline in August.

The apparent lack of a "learning curve" at the beginning of data collection may show that the training before interviewing started was adequate, and that the mistakes made may not have been preventable with further training. The large drop in the error rate at the end of the study may be due to the decreased workload. As fewer interviewers were needed, only the best ones were kept on.

7.2 Item Non-Response

Despite the best efforts of the data collection staff, there will be missing data for any study. While unit non-response for High School and Beyond rounds continues to be adjusted for by weighting, this approach is impractical for item non-response. Therefore, an attempt to reduce item non-response was made for the Fourth Follow-Up.

In previous rounds, interviews were conducted by self-administered questionnaires (SAQ's). Unfortunately, respondents often skipped questions incorrectly or gave unrecognizable answers. Therefore, there was more missing data than could have been achieved through personal interviewing. Also, it was often the case that the reason a particular answer was "missing" was unknown. Possible reasons could be refusals, "don't knows" responses, and unintentional skipping.

In the fourth follow-up, interviewing was conducted using Computer Assisted Telephone Interviewing (CATI). This method uses a computer program to guide the interviewer and respondent through the questionnaire, skipping questions as appropriate, thus speeding up the interview. Unlike SAQs, CATI interviewing virtually eliminates missing data attributable to improperly skipped questions.

Twenty-five items were selected for a comparison between third and fourth follow-up data. Refusal and don't know responses were considered to be missing, but legitimate skips were not. Table 7.2 below shows the number of cases of each type of missing data for the 25 selected items for the third follow-up. Table 7.3 does the same for the fourth follow-up.

Table 7.2--The numbers and percentages of certain types of missing responses for each of 25 third follow-up items (N=13,425)

	Unspecified missing	Multiple response	Uncodable verbatim	Refusal	Don't know	Total
Race	199	0	0	0	0	199
(Percentage)	1.48%	0.00%	0.00%	0.00%	0.00%	1.48%
Working for pay	41	0	0	1	0	42
(Percentage)	0.31%	0.00%	0.00%	0.01%	0.00%	0.31%
Spouse in household?	408	0	0	0	0	408
(Percentage)	3.04%	0.00%	0.00%	0.00%	0.00%	3.04%
Applied grad/prof inst.	852	0	0	0	0	852
(Percentage)	6.35%	0.00%	0.00%	0.00%	0.00%	6.35%
Took GRE	282	0	2	3	0	287
(Percentage)	2.10%	0.00%	0.01%	0.02%	0.00%	2.14%
Educational loans?	490	1	0	0	0	491
(Percentage)	3.65%	0.01%	0.00%	0.00%	0.00%	3.66%
How far, schooling?	119	6	3	0	7	135
(Percentage)	0.89%	0.04%	0.02%	0.00%	0.05%	1.01%
Wages, salaries ('84)	507	0	46	234	238	1025
(Percentage)	3.78%	0.00%	0.34%	1.74%	1.77%	7.64%
Wages, salaries ('85)	513	0	47	237	249	1046
(Percentage)	3.82%	0.00%	0.35%	1.77%	1.85%	7.79%
Employment status 3/84	332	0	0	0	0	332
(Percentage)	2.47%	0.00%	0.00%	0.00%	0.00%	2.47%
Employment status 7/86	332	0	0	0	0	332
(Percentage)	2.47%	0.00%	0.00%	0.00%	0.00%	2.47%
Rec'd formal job trng	737	0	2	2	0	741
(Percentage)	5.49%	0.00%	0.01%	0.01%	0.00%	5.52%
Jobs/trng different?	1129	0	0	0	0	1129
(Percentage)	8.41%	0.00%	0.00%	0.00%	0.00%	8.41%
Gotten job w/o trng?	1129	0	0	0	0	1129
(Percentage)	8.41%	0.00%	0.00%	0.00%	0.00%	8.41%
Satis. w/ supervisor?	991	16	0	0	0	1007
(Percentage)	7.38%	0.12%	0.00%	0.00%	0.00%	7.50%
Satis. co-worker relat	919	1	0	0	0	920
(Percentage)	6.85%	0.01%	0.00%	0.00%	0.00%	6.85%
Marital status, 2/86	75	0	4	4	0	83
(Percentage)	0.56%	0.00%	0.03%	0.03%	0.00%	0.62%
Success in work impt	625	0	0	0	0	625
(Percentage)	4.66%	0.00%	0.00%	0.00%	0.00%	4.66%
Better opp. 4 kids imp	665	3	0	0	0	668
(Percentage)	4.95%	0.02%	0.00%	0.00%	0.00%	4.98%
Volun. Union, etc.	717	4	0	0	0	721
(Percentage)	5.34%	0.03%	0.00%	0.00%	0.00%	5.37%

Table 7.2--The numbers and percentages of certain types of missing responses for each of 25 third follow-up items (N=13,425) (continued)

	Unspecified missing	Multiple response	Uncodable verbatim	Refusal	Don't know	Total
Regis. to vote?	619	3	0	0	0	622
(Percentage)	4.61%	0.02%	0.00%	0.00%	0.00%	4.63%
No. of children	81	0	2	5	0	88
(Percentage)	0.60%	0.00%	0.01%	0.04%	0.00%	0.66%
First inst. type	254	7	0	0	0	261
(Percentage)	1.89%	0.05%	0.00%	0.00%	0.00%	1.94%
First inst. 1st month	180	0	17	0	2	199
(Percentage)	1.34%	0.00%	0.13%	0.00%	0.01%	1.48%
First inst. degree?	429	0	0	0	0	429
(Percentage)	3.20%	0.00%	0.00%	0.00%	0.00%	3.20%
Average	505.0	1.6	4.9	19.4	19.8	550.8
(Percentage)	4.00%	0.01%	0.04%	0.15%	0.16%	4.36%

Table 7.3--The numbers and percentages of certain types of missing responses for each of 25 fourth follow-up items (N= 2,640)

	Unspecified missing	Multiple response	Uncodable verbatim	Refusal	Don't know	Total
Race	6	0	0	50	8	64
(Percentage)	0.05%	0.00%	0.00%	0.40%	0.06%	0.51%
Working for pay	4	0	0	0	0	4
(Percentage)	0.03%	0.00%	0.00%	0.00%	0.00%	0.03%
Spouse/partn. in hshld	21	0	0	1	0	22
(Percentage)	0.17%	0.00%	0.00%	0.01%	0.00%	0.17%
Applied grad/prof inst.	36	0	0	155	66	257
(Percentage)	0.28%	0.00%	0.00%	1.23%	0.52%	2.03%
Took GRE	0	0	0	0	0	0
(Percentage)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Educational loans?	13	0	0	63	50	126
(Percentage)	0.10%	0.00%	0.00%	0.50%	0.40%	1.00%
How far, schooling?	206	0	0	0	0	206
(Percentage)	1.63%	0.00%	0.00%	0.00%	0.00%	1.63%
Wages, salaries ('90)	518	0	12	183	377	1090
(Percentage)	4.10%	0.00%	0.09%	1.45%	2.98%	8.62%
Wages, salaries ('91)	505	0	16	180	293	994
(Percentage)	4.00%	0.00%	0.13%	1.42%	2.32%	7.86%
Employment status 9/89	1	0	0	6	40	47
(Percentage)	0.01%	0.00%	0.00%	0.05%	0.32%	0.37%
Employment status 1/92	1	0	0	7	34	42
(Percentage)	0.01%	0.00%	0.00%	0.06%	0.27%	0.33%
Rec'd formal job trng	0	0	0	51	38	89
(Percentage)	0.00%	0.00%	0.00%	0.40%	0.30%	0.70%
Jobs/trng different?	2	0	0	50	113	165
(Percentage)	0.02%	0.00%	0.00%	0.40%	0.89%	1.31%
Gotten job w/o trng?	0	0	0	47	95	142
(Percentage)	0.00%	0.00%	0.00%	0.37%	0.75%	1.12%
Satis. w/ supervisor?	0	0	0	134	250	384
(Percentage)	0.00%	0.00%	0.00%	1.06%	1.98%	3.04%
Satis. co-worker relat	0	0	0	97	153	250
(Percentage)	0.00%	0.00%	0.00%	0.77%	1.21%	1.98%
Marital status, 1/92	154	0	0	6	11	171
(Percentage)	1.22%	0.00%	0.00%	0.05%	0.09%	1.35%
Success in work impt	2	0	0	53	59	114
(Percentage)	0.02%	0.00%	0.00%	0.42%	0.47%	0.90%
Better opp. 4 kids imp	2	0	0	74	129	205
(Percentage)	0.02%	0.00%	0.00%	0.59%	1.02%	1.62%

Table 7.3--The numbers and percentages of certain types of missing responses for each of 25 fourth follow-up items (N= 2,640) (continued)

	Unspecified missing	Multiple response	Uncodable verbatim	Refusal	Don't know	Total
Volun. union, etc.	5	0	0	0	0	5
(Percentage)	0.04%	0.00%	0.00%	0.00%	0.00%	0.04%
Regis. to vote?	4	0	0	46	84	134
(Percentage)	0.03%	0.00%	0.00%	0.36%	0.66%	1.06%
No. of children	0	0	0	0	0	0
(Percentage)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
First inst. type	490	0	0	1	0	491
(Percentage)	3.88%	0.00%	0.00%	0.01%	0.00%	3.88%
First inst. 1st month	67	0	0	0	0	67
(Percentage)	0.53%	0.00%	0.00%	0.00%	0.00%	0.53%
First inst. degree?	866	0	0	0	0	866
(Percentage)	6.85%	0.00%	0.00%	0.00%	0.00%	6.85%
Average	116.1	0.0	1.1	48.2	72.0	237.4
(Percentage)	0.92%	0.00%	0.01%	0.38%	0.57%	1.88%

Looking first at the overall picture, we can see that for these 25 items, the percentage of missing items drops from over 4% overall to under 2% (4.36% to 1.88%), a reduction of 56.9%. We also see that we have eliminated a whole category of missing data, multiple responses, and have uncodable verbatim's for only the two income variables. Furthermore, we know more about the missing data in the fourth follow-up. In the third follow-up, only 7.2% of the missing data is classified as refusals or don't knows. In the fourth follow-up, 50.9% of the missing data is classified as refusals or don't knows.

We can formally test if there is less missing data in the fourth follow-up, item by item. First, we treat whether or not the item is missing for each respondent as a binary variable, missing or not. Those respondents who have a missing answer on both or neither of the questionnaires tell us nothing about the relative rates of missing data between the two questionnaires. Therefore, our analysis only includes those respondents (different for each item pair) who have a missing response on exactly one of the two questionnaires. If the two items have equal percentages of missing data, we would expect half of the respondents for each item pair to have a missing response on each of the two questionnaires. Therefore, the test is a simple binomial test of whether the percentage of respondents with a missing response on the fourth follow-up is 50%. The results are shown below in Table 7.4. One test is done for each item pair.

Table 7.4--A comparison of numbers of fourth follow-up missing values to numbers of third follow-up missing values

	Number "missing" but not missing in other round		
	3rd FU	4th FU	t-value
Race	86	56	-2.52*
Working for pay	42	4	-5.60**
Lived w/ spouse	329	22	-16.39**
Applied to grad/prof school	239	213	-1.22
Took GRE	98	0	-9.90**
Took out loans for education	325	50	-14.20**
Highest degree planned	107	142	2.22*
Salary, 1 yr before interview	705	852	3.73**
Salary, most recent year	727	767	1.03
Unemployment status, 16 mo. ago	256	16	-14.55**
Current unemployment status	251	11	-14.83**
Rec'd formal training for job	615	73	-20.66**
Job is diff. from training	954	27	-29.60**
Could've gotten job w/o training	956	24	-29.77**
Satisfied with supervisor	801	286	-15.62**
Satisfied with co-worker relations	751	191	-18.25**
Current marital status	62	13	-5.66**
Success in line of work impt	501	91	-16.85**
Better opp for children impt	533	175	-13.45**
Member of union,trade,farm assoc.	586	0	-24.21**
Registered to vote	496	112	-15.57**
Number of children	72	0	-8.49**
1st PSE institution type	159	141	-1.04
1st PSE inst. month started	106	12	-8.65**
Degree for 1st school?	266	320	2.23*

* Significant at .05 ("Significant")
** Significant at .001 ("Very significant")

Variables with less missing data in FU4:	18	"Very significant"
	1	"Significant"
Variables with less missing data in FU3:	1	"Very significant"
	2	"Significant"
Variables with no significant difference:	3	

	25	

The fact that most of the 25 tests show a "very significant" decline in missing data from the third follow-up to the fourth supports our contention that missing data has been reduced in the fourth follow-up of High School and Beyond.

7.3 Consistency Between Third Follow-up and Fourth Follow-up Responses

For the following analysis, we selected only those respondents who completed both the third and fourth follow-up instruments. This sub-population of students will be referred to as the "joint respondents" throughout this section. Theoretically, both answers for these joint respondents should be the same.

In this section, we will look at two items to see how consistent the responses for joint respondents are. One of these items, race/ethnicity, is a variable which should not change over time. However, it does have some definitional problems, as will be seen below. The other item, marital status, should be consistent.

7.3.1 Race/Ethnicity

Race/ethnicity is a characteristic of the respondents that should not change between the third and fourth follow-up surveys. Since we have independent answers to race/ethnicity from the two surveys, we compare the two answers below. A complete cross-tabulation is given below:

Table 7.5--A cross-tabulation of the joint respondents' responses to the race/ethnicity questions on each survey*

RACE: FU4	Hispanic	Native American	Asian/ Pacific	Black	White	Row totals
FU3						
Hispanic	1529	21	10	115	264	1939
Native American	10	147	1	26	91	275
Asian/Pacific	8	5	301	8	37	359
Black	10	2	2	1657	35	1706
White	67	34	8	9	7912	8030
Column totals	1624	209	322	1815	8339	12309

* Cases that were classified as missing or unknown in either follow-up are excluded, because they do not indicate a discrepant response.

If the two surveys were to match exactly, all of the off-diagonal entries in the above table would be zero. The best way to summarize this data would be to see what percentage of cases match. The next table below shows what percentage of the joint respondents gave the same answer in the fourth follow-up that they gave in the third follow-up, separated by how they responded in the third follow-up. It also shows what percentages gave each of the possible "non-matching" answers in the fourth follow-up.

Table 7.6--Third follow-up race responses compared to fourth follow-up responses

RACE: FU4	MATCH	NON-MATCHES Hispanic	Native American	Asian/ Pacific	Black	White
FU3						
Hispanic	78.9%	-	1.1%	0.5%	5.9%	13.6%
Native American	53.5%	3.6%	-	0.4%	9.5%	33.1%
Asian/Pacific	83.8%	2.2%	1.4%	-	2.2%	10.3%
Black	97.1%	0.6%	0.1%	0.1%	-	2.1%
White	98.5%	0.8%	0.4%	0.1%	0.1%	-
Overall	93.8%					

Overall, of the 12,309 respondents who gave their ethnicity on both questionnaires, 11,546 (93.8%) gave the same ethnicity on both. However, certain race/ethnicity categories (e.g., Native American) have substantially less agreement. Only 53.45% of the joint respondents who classified themselves as Native American during the third follow-up classified themselves as Native Americans again during the fourth follow-up. The table above illustrates that when mon-matching response is given, the answer tends to be "white."

One explanation may be that the method of administering the question changed between rounds. Unlike the third follow-up, which involved self-administered questionnaires, the fourth follow-up was done by telephone. The questionnaires mailed during the third follow-up had the five race/ethnicity categories listed. However, over the telephone, respondents who were simply asked, "What is your race/ethnicity?" Then, their answers were coded by the interviewers. It is possible that Native Americans, Hispanics, and Asian/Pacific Islanders to classify themselves as Black or White, not knowing that there was a more specific category for them, thus leading to more Blacks and Whites in the fourth follow-up.

7.3.2 Marital Status

In the third follow-up, respondents were asked about their marital status in the first week of February, 1986. In the fourth follow-up, respondents were asked about their marital status during and since February, 1986. Therefore, we again have two answers to marital status during February, 1986. [One note of caution, however, is that the respondents were asked about the first week of 1986 in the third follow-up, but no particular week of February was specified in the fourth follow-up. Therefore, any respondents who had a change in marital status during the last three weeks of February, 1986, could give differing answers.

The proportion of cases in which this could have occurred is probably small.] The data are given below:

Table 7.7--A cross-tabulation of the february, 1986 marital status of the joint respondents, as reported on the third and fourth follow-ups.

FU4 February, 1986.	Never married Relat.	Divorced Totals	Widowed	Separ.	Married	Marr.- like rstatus	Row back in
FU3							
Never Married	8157	6	1	7	147	4	8322
Divorced	9	188	0	2	12	0	211
Widowed	0	0	6	0	3	0	9
Separated	15	23	0	134	33	1	206
Married	62	27	1	15	2445	3	2553
Marr.-like Rel.	127	5	1	2	37	381	553
Column Totals	8370	249	9	160	2677	389	11854

Again, if the two surveys were to match exactly, all of the off-diagonal entries in the above table would be zero. The best way to summarize this data is again to see what percentage of cases match. The next table below shows what percentage of the joint respondents gave the same answer in the fourth follow-up that they gave in the third follow-up. It also shows the percentages of the "non-matching" answers in the fourth follow-up. Those respondents with a missing response for either questionnaire were excluded from the percentages below.

Table 7.8--Third follow-up marital status responses compared to fourth follow-up responses

		Non-matches					
	FU4 matches	Never married	Divorced	Widowed	Separ.	Married	Marr.- like relat.
FU3							
Never married	98.0%	-	0.1%	0.0%	0.1%	1.8%	0.1%
Divorced	89.1%	4.3%	-	0.0%	1.0%	5.7%	0.0%
Widowed	66.7%	0.0%	0.0%	-	0.0%	33.3%	0.0%
Separated	65.0%	7.3%	11.2%	0.0%	-	16.0%	0.5%
Married	95.8%	2.4%	1.1%	0.0%	0.6%	-	0.1%
Marr.-like rel.	68.9%	23.0%	0.9%	0.2%	0.4%	6.7%	-
Overall	95.4%						

Overall, of the 11,854 respondents who gave their marital status on both questionnaires, 11,311 (95.4%) had answers that agreed. Unlike the race/ethnicity question, memory and timing play an important role in matching answers for marital status. In this case, the recall period for third follow-up respondents was years shorter than the recall period for fourth follow-up respondents. After all, respondents were asked in 1986 about a relatively recent event, while in 1992, they were asked to recall their status back in February, 1986.

BEST COPY AVAILABLE

As with the race/ethnicity question, method of administering the question differed between rounds: the question formatting had changed and the fourth follow-up used preloaded data to verify status.

7.4 Proprietary Institution Non-response Issues

Proprietary (i.e., private, for profit) institutions had a much higher non-response rate than other types of institutions. In this section, we will look at non-response and student characteristics. In order to evaluate the potential for bias, we will compare respondents to non-respondents among proprietary school students. Next, we will compare proprietary school students to two other groups: non-proprietary school students and students who attended both proprietary and non-proprietary institutions. The comparisons are made on three demographic variables: race/ethnicity, socio-economic status, and gender.

7.4.1 Proprietary Respondents vs. Proprietary Non-respondents

Table 7.9 shows a slightly higher response rate among whites, but all rates ranged between 50% and 65%. A chi-square test of independence was not significant at the .05 level (Chi-square = 9.09, df = 4, p = 0.058).

Table 7.9--Race/ethnicity profiles, by response categories

Race/Ethnicity:	Hispanic	Native American	Asian/ Pacific	Black	White	Total
		Number				
Non-respondents	86	8	12	88	189	383
Respondents	86	12	12	118	313	541
Total	172	20	24	206	502	924
		Percentage				
Non-respondents	22.5%	2.1%	3.3%	23.0%	49.4%	100.0%
Respondents	15.9%	2.2%	2.2%	21.8%	57.9%	100.0%
Response rate	50.0%	60.0%	50.0%	57.3%	62.4%	58.6%

Table 7.10 indicates that the non-respondents are not systematically different from the respondents in socio-economic status. This was confirmed by a chi-square test of independence (Chi-square = 1.52, df = 3, p = 0.67).

Table 7.10--Socio-economic status quartiles, by response categories

Socio-economic status	Lowest quartile	Second quartile	Third quartile	Highest quartile	Total
		Number			
Non-respondents	126	90	104	53	373
Respondents	161	143	147	79	530
Total	287	233	251	132	903
		Percentage			
Non-respondents	33.8%	24.1%	27.9%	14.2%	100.0%
Respondents	30.4%	27.0%	27.7%	14.9%	100.0%
Response rate	56.1%	61.4%	58.6%	59.9%	100.0%

Table 7.11 suggests that the proprietary school non-response rate for females is higher than for males. This is confirmed by a chi-square test of independence (Chi-2 = 14.70, df = 1, $p < 0.001$).

Table 7.11--Gender profiles, by response categories

Gender	Male	Female	Total
	Number		
Non-respondents	116	268	384
Respondents	232	314	546
Total	348	582	930
	Percentage		
Non-respondents	30.2%	69.8%	100.0%
Respondents	42.5%	57.5%	100.0%
Response rate	66.7%	54.0%	58.7%

7.4.2 Proprietary School Students vs. Non-proprietary School Students

We have seen that females were much more likely than males to attend non-respondent proprietary schools, but that there were no other significant differences among the other two demographic categories: race/ethnicity and socio-economic status. We will now assess how students who attended proprietary institutions are different from students who attended other types of institutions. In order to do this, we first classified students into three categories: proprietary institution students only, students who have attended a mix of proprietary and non-proprietary institutions, and non-proprietary students. First, are the students who attended only proprietary institutions different from those who attended a mix of proprietary and one non-proprietary institution? Second, are the students who attended at least one proprietary institution different from those who only attended non-proprietary institutions?

Table 7.12 illustrates some clear differences in three groups. Hispanics, Native Americans, and Blacks are much more likely to go to proprietary institutions, while whites and Asian/Pacific Islanders are more likely to go to the non-proprietary institutions. In fact, significant differences were found between the proprietary only and Proprietary/Non-proprietary group (Chi-2 = 12.717, df = 4, p = 0.012), and between proprietary only and non-proprietary only students (Chi-2 = 65.767, df = 4, p<0.001).

Table 7.12--A comparison of the race/ethnicity by institution attended

Race/Ethnicity	Hispanic	Native American	Asian/ Pacific	Black	White	Total
			Number			
Proprietary only	110	13	6	123	316	568
Prop. and non-pr	110	13	29	125	318	595
Non-pr only	1213	156	354	1122	5494	8339
Total	1433	182	389	1370	6128	9502
			Percentage			
Proprietary only	7.7%	7.1%	1.5%	9.0%	5.2%	6.0%
Prop. and non-Pr	7.7%	7.1%	7.5%	9.1%	5.2%	6.3%
Non-pr only	84.7%	85.7%	91.0%	81.9%	89.7%	87.8%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 7.13 illustrates the clearest difference among the institution attenders. As socio-economic status increases, so does the chance that the respondent will have gone to non-proprietary institutions exclusively. As socio-economic status decreases, the chance that the respondent will have gone to a proprietary institution increases. Significant differences were found between the proprietary/non-proprietary institutions vs. proprietary-only (Chi-2 = 52.84, df = 3, p<0.001 only) and between the non-proprietary only vs. proprietary (Chi-2 = 181.79, df = 3, p<0.001).

Table 7.13--A comparison of the socio-economic status by institution type

Socio-economic status	Lowest quartile	Second quartile	Third quartile	Highest quartile	Total
			Number		
Proprietary only	213	149	141	44	547
Prop. and non-pr	148	138	178	124	588
Non-pr only	1547	1738	2184	2815	8284
Total	1908	2025	2503	2983	9419
			Percentage		
Proprietary only	11.2%	7.4%	5.6%	1.5%	5.8%
Prop. and non-pr	7.8%	6.8%	7.1%	4.2%	6.2%
Non-pr only	81.1%	85.8%	87.3%	94.4%	88.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

Table 7.14 shows there does not seem to be a difference between the proprietary-only and proprietary/non-proprietary groups with respect to gender. This is confirmed by a chi-square test (Chi-2 = 0.556, df = 1, p=0.46). However, females are more likely to attend at least one proprietary institution (Chi-2 = 44.16, df=1, p<0.001).

Table 7.14--A comparison of the gender by institution type

Gender	Male	Female	Total
		Number	
Proprietary only	215	358	573
Prop. and non-pr.	212	385	597
Non-pr only	4000	4382	8382
Total	4427	5125	9552
		Percentage	
Proprietary	4.9%	7.0%	6.0%
Prop. and non-pr	4.8%	7.5%	6.3%
Non-pr only	90.4%	85.5%	87.8%
Total	100.0%	100.0%	100.0%

Appendix A:

Transcript Request Packages

NATIONAL LONGITUDINAL STUDIES PROGRAM

High School and Beyond
A National Longitudinal Study for the 1980's

Sponsored by the Center for Education Statistics,
U.S. Department of Education

The professional organizations listed below fully endorse the Postsecondary Education Transcript Study and encourage their members to cooperate in this important project.

American Association of Collegiate Registrars and Admissions Officers (AACRAO)

American Association of Community Colleges (AACC)

American Association of State Colleges and Universities (AASCU)

American Council on Education (ACE)

Council of Graduate Schools (CGS)

National Association of Student Financial Aid Administrators (NASFAA)

National Institute of Independent Colleges and Universities (NIICU)

February 1993

Dear Registrar:

NORC, a social science research center at the University of Chicago, requests your assistance in the conduct of a Postsecondary Education Transcript Study. We seek your help in collecting transcripts for a sample of students who are participating in the High School and Beyond Survey (HS&B:92) sponsored by the National Center for Education Statistics (NCES). The purpose of the transcript study, a component of HS&B:92, is to obtain reliable and objective information about the types and patterns of course-taking patterns to student characteristics available in student questionnaire files, and to subsequent occupational choice and success.

In 1992 the National Opinion Research Center (NORC) at the University of Chicago, under the sponsorship of the U.S. Department of Education National Center for Education Statistics (NCES), surveyed 14,000 members of the high school sophomore class of 1980 using computer-assisted telephone interviewing. This, the fourth follow-up to the study High School and Beyond (HS&B), will mark the fifth time that NORC has surveyed this population. HS&B began in 1980, and this latest data collection interviewed the sample members when they were 10 years out of high school. HS&B has proved to be one of the most valuable longitudinal studies conducted by the Department of Education based upon the large volume of research that has used its rich data files. The project is conducted under the guidance of Dr. C. Dennis Carroll, who is the Chief of the Longitudinal Studies Branch of the NCES Postsecondary Education Statistics Division.

We would like to obtain the transcripts of one or more sample members who reported attending your school. Specifically we are requesting photocopies of transcripts for each individual named on the enclosed checklist for the years reported by the student for his or her attendence. We would also appreciate it if you could provide us with: 1) a copy of the school's course catalog and 2) an interpretation of your grading system in order to facilitate accurate and uniform coding of the data.

Privacy and confidentiality are always of concern to institutions and offices that maintain student records. NCES and the organizations under contract to it adhere to the highest standards in protecting the privacy of individuals involved in the research it undertakes. Appropriate measures are employed to ensure the confidentiality of research participants during the collection, analysis, and reporting of all survey data. Of course, all relevant safeguards will be applied to this study.

As in the past, survey data are being collected under the provision of the Family Education Rights and Privacy Act (FERPA) that allows the release of records to the Secretary of Education or his agent without prior written consent by survey subjects.

Endorsement of the transcript study has been made by the American Association of Collegiate Registrars and Admissions Officers. A copy of the article endorsing the study is included in this folder.

We would appreciate return of the requested materials by March 5, or as soon thereafter as possible. Reimbursement for all transcripts will be made if you request it, and a voucher has been included for this purpose.

If we can assist you in any way to provide these materials, or if you have any questions about the study, please do not hesitate to call Dr. C. Dennis Carroll, Branch Chief Officer, Transcript Study at (202) 219-1774 (collect) or Patricia Marnell, Transcript Study Project Manager, (312) 753-7823.

Sincerely,

Barbara K. Campbell, Ph.D.
High School and Beyond
Project Director

BKC/rlp

December 1992/January 1993 Newsletter of
American Association of Collegiate Registrars and Admissions Officers

Transcript Alert

Transcripts will be collected for the High School and Beyond study (HS&B) in January and February 1993. The project is being conducted by the National Opinion Research Center (NORC) at the University of Chicago for the National Center for Education Statistics (NCES) of the U.S. Department of Education. NCES and NORC have been working with AACRAO on the project since 1986. Institutions will be reimbursed for supplying the transcripts as necessary. NCES guidelines and Congressional legislation mandate strict confidentiality requirements for the study, to which NORC adheres. The Family Education Rights and Privacy Act (FERPA) grants permission for NCES studies to collect the transcripts and the new Higher Education Amendments make participation no longer voluntary. The data from transcripts collected on other studies like HS&B have proved to be very valuable for policy makers and researchers analyzing patterns in course taking and eventual labor market success. We encourage your expedient handling of the NORC requests.

Dear Registrars and Officials:

As part of its Longitudinal Studies program, the National Center for Education Statistics has been collecting transcript and other information for persons who have participated in its surveys. To continue this effort, the Center has authorized the National Opinion Research Center (NORC) to obtain student transcript data for individuals who are participating in the High School and Beyond (HS&B) survey. The goal of this study is to provide information which can be aggregated to examine research issues at the national level. Education researchers and policy analysts will relate the information about courses taken and credits earned to the characteristics gathered from questionnaires and other sources. HS&B will enable researchers to analyze the relationships between course taking patterns, academic achievement, and subsequent occupational choices and success. Student names are used only to make sure that data on variables from different sources (test, questionnaires, and transcripts) refer to the same individuals and not to find out anything about particular individuals.

The grant of authority for collection of the transcript data is made pursuant to the provision in the Family Education Rights and Privacy Act (FERPA), implemented by ???, that allows the release of records to the Secretary of Education or to his agent without the prior consent of the survey participants. The privacy of the information you are asked to supply to NORC will be protected, as requiredd by FERPA. A copy of the relevant section of the act is reproduced on the reverse side of this page.

We would appreciate your cooperation with NORC in the transcript study.

. . . .

. . . .

Sincerely yours,

Emerson J. Elliott
Commissioner

EJE/rlp

NORC
National Center For Education Statistics
National Longitudinal Studies Program
High School and Beyond

NCES's Longitudinal Studies Program

The mandate of the National Center for Education Statistics (NCES) of the U.S. Department of Education includes the responsibility to "collect and disseminate statistics and other data related to education in the United States" and to "conduct and publish reports on specific analyses of the meaning and significance of such statistics" (Education Amendments of 1974 - Public Law 93-380, Title V, Section 501, amending Part A of the General Education Provisions Act).

Consistent with this mandate and in response to the need for policy-relevant, time-series data on a nationally representative sample of high school students, NCES instituted the National Longitudinal Studies (NLS) program, a continuing long-term project. The general aim and personal development of high school students and the personal, familial, social institutional, and cultural factors that may affect that development.

The NLS program was planned to make use of time-series databases in two ways: (1) each cohort is surveyed at regular intervals over a span of years, and (2) comparable data is obtained from successive cohorts, permitting studies of trends relevant to educational and career development and societal roles. High School and Beyond (HS&B) is a major study in the NLS program.

High School and Beyond

High School and Beyond (HS&B) is a longitudinal study of the critical transition years as high school students leave the secondary school system to begin postsecondary education, work, and family formation. Its purpose is to provide information on the characteristics, achievements, and plans of high school students, their progress through high school, and the transition they make from high school to adult roles. Because of the breadth of the survey's coverage, data can be used to examine such policy issues as school effects, bilingual education, dropouts, vocational education, academic growth, access to postsecondary education, student financial aid, and life goals. High School and Beyond was designed to collect data that would be comparable to that of the National Longitudinal Study of the High School Class of 1972 (NLS-72).

In 1980, a national sample of over 30,000 sophomores and 28,000 seniors enrolled in 1,015 public and private schools participated in the Base Year Survey. During this stage of the study, students completed a cognitive test and a questionnaire about their high school experiences and plans for the future. In order to find out how plans have worked out or changed, subsamples of the base-year students were asked to complete follow-up questionnaires in 1982, 1984, 1986 and 1992. The 1980 sophomore class also completed a cognitive test in 1982 when they were seniors. In addition, base-year data were compiled from such sources as school administrators, teachers, students' administrative records (transcripts), and parents of selected students.

In the spring of 1984 a consortium of university research centers sponsored a study of principals; guidance, vocational, and community service program counselors; and up to 30 teachers in each on of a sample of approximately 500 HS&B schools. Results of this survey, funded by the National Institute of Education, have become part of the HS&B database and permit researchers to describe the impact of the school environment on the educational process.

Postsecondary transcripts were collected for the senior cohort of HS&B in 1984. They contain reliable and objective information about the types and patterns of courses taken by students in colleges, graduate schools, and non-collegiate postsecondary institutions. The information has been merged with the expanding HS&B database. It will be possible for researchers to relate course-taking patterns to student characteristics available in the student questionnaire data files and to subsequent occupational choice and success.

A Financial Aid Records Study was conducted in 1985 for the senior cohort and in 1987 for the sophomore cohort. Postsecondary schools attended by HS&B students provided data on the students' costs of attendance, student and family contributions, and financial aid packages. Guaranteed Student Loan records and Pell Grant information were collected from central data bases maintained in the Office of Education. Data from the three sources were then merged to provide a comprehensive profile of financial assistance.

In 1986 records were requested for Guaranteed Student Loans and Pell Grants that HS&B sophomores may have obtained. This financial aid information was collected to complement the postsecondary education transcripts. A survey of the 1980 sophomore cohort's postsecondary transcripts was conducted in 1987. Some 3,100 postsecondary institutions were asked to participate in this study.

In 1992 a Computer Assisted Telephone Interview (CATI) was conducted with the 1980 sophomore cohort. A postsecondary transcript survey is also underway for this cohort.

Hence, for the 1980 sophomore class, the Department of Education will have a complete record of high school experiences and past high school activities, including postsecondary schooling and financing. Like that of the senior cohort, the patterns of courses taken by students will allow researchers to relate course-taking patterns to student characteristics available in the student questionnaire data files, and to subsequent occupational choice and success.

High School and Beyond Fourth Follow-Up, Sophomore Cohort (HS&B:92)

INSTRUCTIONS

Participation in the Postsecondary Education Transcript Study involves obtaining transcripts and related materials from your files and sending them to NORC, a social science research center at the University of Chicago. The steps on the following pages provide details on:

Step 1: Review student checklist

The Student checklist provides the names, in alphabetical order, of the student for whom copies of the transcript are being requested. In addition, other names (e.g., maiden, family, alternate spelling, etc.), social security numbers, and birthdates are provided as additional identifying information for many students. Please enter a mark if you are enclosing a transcript(s) for a student. If you are unable to provide some or any records for a student, please check either "No Record of Student," "Completed No Courses" or indicate another reason in the space provided.

EXAMPLES:

"Never attended this school"

"Transcripts cannot be located at this time"

"Did not attend long enough to earn credit"

Two copies of the student checklist have been enclosed. Please return one copy with your checkmarks and any comments with the transcripts. The other copy is for your school's records.

Step 2: Retrieve and prepare transcripts

Locate and prepare (e.g., photocopy, generate a computer printout, etc.) a copy of each transcript for each student on the checklist.

Step 3: Label the transcripts

Affix the enclosed student labels to the BACK of the appropriate transcripts.

Step 4: Insert disclosure notices in each student's record file

Disclosure notices indicating the purpose for which student records were accessed for the transcript study are enclosed for your convenience.

Step 5: Obtain course catalog(s) or course list(s)

Obtain course catalog(s) or course list(s) describing the courses offered by your insititution. Catalogs should be included for all programs and schools for which the student has been enrolled (e.g., the liberal arts college and the law school). Please indicate on the checklist whether the current catalog(s) or course list(s) has been included in the package for return to NORC.

Step 6: Obtain grading system description

Obtain a copy of your school's official description of its grading system and/or other method of evaluating student performance. This might include, for example, an explanation of the meaning of letter grades (e.g., A,B...F), non-letter grading (e.g., Pass, High-Pass, Honors, etc.), and /or other standard codes for the evaluation of student performance. In many instances, this would entail translation of grade designations to verbal (e.g., an "A" = ("Outstanding work"), or quantitative (e.g., "A" = "95-100") definitions.

Step 7: For reimbursement of expenses

If you would like to be reimbursed for the photocopying required for the transcripts or for other related expenses, please complete and return all copies of the enclosed voucher with the transcripts. One copy of the voucher will be returned with a check that will be issued upon receipt of the transcript package. If you have any questions regarding reimbursement, please call Patricia Marnell, Transcript Study Project Manager, at (312) 753-7823.

Step 8: Assemble and send transcripts to NORC

A pre-paid, business reply label is enclosed for returning the transcripts and other related materials. Please use the enclosed return address label with your institution's name, mailing address and identifying bar code. This will aid NORC in receipting your package more quickly. These labels are in the right-hand flap of this folder.

Please return all transcript study materials by March 5. If you encounter problems of any kind in regard to our request for transcript, or you are unable to mail them by March 5 or shortly thereafter, please call Patricia Marnell, Transcript Study Project Manager, at (312) 753-7823.

United States
Department of Education
Washington, DC 20208–5654

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Fourth Class Special
Special Handling